



# Variation du nombre de copies de plasmides au sein populations monoclonales de bacteries

Jérôme Wong Ng

## ► To cite this version:

Jérôme Wong Ng. Variation du nombre de copies de plasmides au sein populations monoclonales de bacteries. Biophysique [physics.bio-ph]. Université Pierre et Marie Curie - Paris VI, 2008. Français. NNT : . tel-00522991

**HAL Id: tel-00522991**

**<https://theses.hal.science/tel-00522991>**

Submitted on 4 Oct 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité Physique des liquides

présentée par

Jérôme WONG NG

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

VARIATION DU NOMBRE DE COPIES DE PLASMIDES  
AU SEIN POPULATIONS MONOCLONALES DE BACTÉRIES

Soutenance prévue le 10 décembre 2008 devant le jury composé de :

Mme.	Sophie Cribier	Examinatrice
M.	Patrick Tabeling	Rapporteur
M.	Bahram Houchmandzadeh	Rapporteur
M.	François Taddei	Examineur
M.	Didier Chatenay	Directeur de thèse
M.	Jérôme Robert	Co-directeur de thèse

---

# Remerciements

Dérogeant à la règle de la D22 d'expédier cette formalité que sont les remerciements, je souhaiterais m'étaler légèrement sur celles-ci, tant il y a de personnes sans qui cette thèse aurait duré 107 ans.

Par contre, je commencerais par remercier la sous-préfecture d'Anthony qui m'aura fait perdre bien des journées, qui m'aura fait passer de longues heures dans le froid plusieurs matin de décembre par an, qui m'aura pourri le reste de ces journées tant j'étais scandalisé mais qui m'aura appris à rester calme (en apparence du moins et surtout face au guichetier) pour faire face à mes futures aventures administratives.

Plus sérieusement, je remercie ceux qui m'ont financé ou qui m'ont permis de l'être.

Je remercie ensuite les secrétaires du laboratoire : Annie pour sa bonne humeur constante (pas comme Nora :P), Marie spécialement pour avoir traité mon dossier de financement, et Nora, "dragon" efficace qui passe les commandes sur on ne sait quel support.

Je remercie les membres du jury d'avoir accepté cette lourde tâche qu'est d'examiner/rapporter une thèse. Plus particulièrement, je remercie François Taddei pour ses encouragements et je souhaite remercier le plus chaleureusement possible Bahram Houchemandzadeh, qui a rapporté cette thèse avec toute l'indépendance et la justesse qui le caractérise.

Je voudrais tenter d'exprimer toute ma gratitude envers Didier. J'ai énormément apprécié et appris de l'intérêt, l'enthousiasme et l'humilité qu'il conserve face à l'océan d'inconnues que comporte la biologie. Je tiens à préciser que ses paroles (sur des sujets personnels ou scientifiques) sont toujours beaucoup plus sensées, profondes et justes que son ton (qu'il soit sérieux, qu'il rigole ou qu'il grogne), sa barbe et sa bonhomie ne le laisseraient entendre. Quand il n'était pas là ou pas accessible, c'était pour discuter avec un post-doctorant pour éventuellement le conseiller sur son dossier CNRS, pour aider un thésard à trouver un post-doc ou encore pour tenter de faire en sorte qu'il existe encore des postes de permanents dans 10 ans. Il est toujours rassurant de savoir que certaines personnes se dévouent autant pour les plus jeunes. Il a toujours été présent aux moments clés (à bien d'autres moments aussi d'ailleurs) et il a été un soutien constant et toujours encourageant malgré trois longues années de données non reproductibles.

Ensuite, je dois remercier Jérôme (Robert, je ne parles pas de moi à la troisième personne). Il a été présent tout au long de cette thèse. Personne ne sait comment il est possible de préparer les milieux de cultures, s'assurer que tous les produits nécessaires sont en stock, passer les commandes nécessaires, réparer la pompe à palette ou encore un de ces trucs qui ne marchent plus vraiment d'équerre, faire un gel en attendant que la PCR en cours arrive à terme, commencer la préparation des bactéries ou la terminer quand je me suis mal organisé, faire la préparation quand je suis pas là, discuter des résultats avec moi, s'étonner de la complexité de la biologie, rechercher encore une explication logique, recommencer ce qu'il a fait la veille et l'avant-veille sans se décourager car c'est un peu comme ça qu'on debug la biomol et pour finir (cette liste ne pouvait être exhaustive mais il faut bien la terminer par quelque chose) faire les 200h d'enseignements qui lui incombent car il est toujours en sur-service. Je ne sais pas vraiment comment il fait, mais il y arrive et cela reste un exemple d'organisation. Encore merci pour tout, je tenterais de faire preuve d'autant de générosité au sens large du terme.

Un petit mot de remerciements pour les deux autres permanents du groupe, Georges (et son mégaphone) pour ses disputes avec Didier et Alexis la gachette pour sa compréhension et son aide sur la lithographie.

Pour terminer sur le groupe, je remercie les "anciens" thésards de la D22 : Luis, Nicolas, Julien, avec une mention spéciale à Séb"Untu". Je remercie aussi Stéphane qui me rappelle qu'on peut toujours voir les choses avec encore moins d'espoir et Juliette, "la directrice de Labo" qui crie sur les stagiaires. Ces deux derniers m'ont soutenu jusque dans les derniers instants. Leur soutien était tel que durant la dernière semaine (avant de rendre ce manuscrit), ils se relayèrent l'une le jour et l'autre la nuit pour m'accompagner dans mes trop nombreuses pauses cafés. Un grand merci à eux deux.

Il me faut aussi remercier toutes les autres personnes qui m'ont encouragé, Pierre (le solitaire du dimanche), Etienne, Brian, Maïre, Fabien et Ghyomm.

1

Finalement, le plus grand des merci à Kristy et Aby qui m'ont permis de rester à peu près équilibré.

---

<sup>1</sup>Si vous trouvez bizarre de ne pas apparaître dans ces remerciements, je vous invite à prendre un stylo et à écrire ce que vous voudriez entendre;) .

---

# Résumé

Les plasmides sont des fragments d'ADN extrachromosomaux largement utilisés en tant qu'outil en biologie moléculaire. Un plasmide code pour sa propre fréquence de réplication et peut être à copies multiples au sein de son hôte (e.g. la bactérie). Il peut aussi apporter à son hôte un avantage sélectif tel que la résistance à un antibiotique, ce qui confère un caractère symbiotique au système hôte-plasmide.

Nous voulons explorer : 1-les variations du nombre de copies de plasmides (PCN) afin de comprendre les interactions que le plasmide peut avoir avec son hôte 2- la manière dont se fait la régulation de ce nombre de copies.

Nous nous sommes fixés quelques plasmides modèles dans lesquels nous avons inséré le gène d'une protéine fluorescente (mOrange). Ces plasmides ont été injectés dans une bactérie contenant le gène d'un autre rapporteur fluorescent (eGFP) sur son chromosome ce qui nous permet d'avoir une référence au sein de chaque bactérie. Nous avons ensuite observé ces bactéries transformées à l'aide d'un dispositif microfluidique développé au laboratoire qui nous permet des observations de bas niveau de fluorescence sur des bactéries isolées.

Pour chaque plasmide étudié, nous avons pu mesurer son PCN moyen par bactérie au sein de la population. Nous avons aussi évalué son PCN moyen à l'aide de techniques usuelles de biologie moléculaire, et avons obtenu approximativement les mêmes résultats.

Nous avons constaté pour certains plasmides des variations d'expression de la mOrange plus grandes que pour la eGFP. Nous attribuons cette augmentation de la variabilité d'expression à la variabilité du PCN. Cette variabilité soit d'autant plus petite que le PCN moyen est grand.

Ensuite, nous avons pu extraire des données la variation du PCN au sein de la population. Par ailleurs, nous donnons aussi à l'aide d'une méthode plus approximative la distribution du PCN par bactérie.

Finalement, nous avons fait croître les bactéries dans un milieu contenant plus ou moins d'antibiotique dont la résistance est portée par le plasmide. Nous n'avons pas réussi à changer les caractéristiques des plasmides, mais avons pu regarder le phénomène de perte dans un de nos plasmides.

# Table des matières

Remerciements	i
Résumé	iii
Table des matières	iii
<b>I Introduction : l'expression génétique chez <i>Escherischia Coli</i></b>	<b>1</b>
<b>1 Le dogme central</b>	<b>3</b>
1.1 La transcription . . . . .	4
1.2 La traduction . . . . .	5
1.3 L'opéron Lactose, un exemple de régulation génétique. . . . .	6
<b>2 Les variations phénotypiques et les réseaux</b>	<b>8</b>
2.1 Variation d'expression d'un gène . . . . .	9
2.2 Variabilité conditionnée de l'expression génétique. . . . .	11
2.3 Variabilité : bruit ou adaptabilité . . . . .	13
2.3.1 Fonction et robustesse . . . . .	13
2.3.2 Choisir la variabilité plutôt qu'un contrôle global . . . . .	14
<b>3 Le chromosome d'<i>E. Coli</i></b>	<b>16</b>
3.1 La réplication du chromosome . . . . .	16
3.1.1 La régulation de la réplication . . . . .	18
3.1.2 La fourche de réplication . . . . .	19
3.2 La partition du chromosome . . . . .	19
<b>4 Les plasmides</b>	<b>21</b>
4.1 Le Plasmide F . . . . .	21
4.1.1 La régulation de la réplication du miniF . . . . .	22
4.1.2 Le système de partitionnement du pZC320 . . . . .	23
4.2 Plasmide R1 . . . . .	23
4.2.1 La régulation de la réplication des plasmides "miniR1par <sup>-</sup> " et "miniR1par <sup>+</sup> " . . . . .	25
4.2.2 Le système de partition du miniR1par <sup>+</sup> . . . . .	26
4.3 Le Plasmide ColE1 . . . . .	27
4.3.1 La régulation du plasmide mini-colE1 . . . . .	27

4.4	Variabilité du nombre de copies de plasmides . . . . .	28
<b>II</b>	<b>Dispositif expérimental</b>	<b>30</b>
<b>5</b>	<b>Les constructions biologiques</b>	<b>31</b>
5.1	L'expression de la protéine fluorescente . . . . .	32
5.1.1	Le promoteur PtacI* . . . . .	32
5.1.2	Le choix des protéines fluorescentes . . . . .	33
5.1.3	La terminaison TR2* . . . . .	33
5.1.4	La construction (double PCR) . . . . .	34
5.2	Une référence sur le chromosome . . . . .	35
5.3	Le plasmide et sa référence . . . . .	37
<b>6</b>	<b>Microscopie</b>	<b>39</b>
6.1	Le $\mu$ FACS . . . . .	40
6.1.1	Cellule microfluidique . . . . .	40
6.1.2	Le microscope inversé . . . . .	40
6.2	Le Timelapse . . . . .	43
<b>III</b>	<b>Résultats</b>	<b>44</b>
<b>7</b>	<b>Une première caractérisation</b>	<b>45</b>
7.1	Temps de division et plasmides . . . . .	45
7.2	Temps d'induction . . . . .	46
7.3	Concentration d'IPTG . . . . .	48
7.4	Variabilité inter colonies et OD d'induction . . . . .	48
<b>8</b>	<b>Estimation du nombre de copies moyen pour chacun des plasmides</b>	<b>50</b>
8.1	PCR en temps réel . . . . .	50
8.2	Calibration des protéines fluorescentes . . . . .	52
8.2.1	Correction de la fluorescence pour les fluorescences moyennes. . . . .	52
8.2.2	Equivalence entre eGFP et mOrange et estimation de la fuite . . . . .	53
8.3	Linéarité de l'expression avec le nombre de copies . . . . .	54
8.4	Expressions moyennes et PCN moyen . . . . .	54
8.5	Effet de la température sur les miniR1 . . . . .	58
<b>9</b>	<b>Variation du nombre de copies de plasmides</b>	<b>59</b>
9.1	Correction de la fluorescence pour les variances . . . . .	59
9.2	Bruit d'expression lié au plasmide . . . . .	60
9.3	Variance du nombre de copies de plasmides . . . . .	61
9.4	Une méthode "graphique" . . . . .	64
9.4.1	Discerner entre plusieurs copies d'un gène . . . . .	64
9.4.2	Construction de la distribution . . . . .	66

---

<b>10 Changement de concentration d'antibiotique</b>	<b>69</b>
10.1 Qualité de la résistance avec le nombre de copies du plasmide . . . . .	69
10.2 Constance du PCN et de sa variabilité . . . . .	70
10.3 Perte du plasmide . . . . .	71
 <b>IV Conclusions</b>	 <b>74</b>
 <b>V Annexes et bibliographie</b>	 <b>77</b>
<b>A Les protéines fluorescentes utilisées</b>	<b>78</b>
<b>B Protocoles</b>	<b>82</b>
B.1 Biologie moléculaire et Microbiologie . . . . .	82
B.1.1 Stocks et Solutions . . . . .	82
B.1.2 Protocoles de biologie moléculaire . . . . .	83
B.1.3 Protocoles détaillés des constructions effectuées . . . . .	84
B.2 Cellules microfluidiques . . . . .	91
B.2.1 Fabrication des moules . . . . .	91
B.2.2 La cellule . . . . .	91
B.2.3 Préparation des bactéries . . . . .	93
 <b>C Analyse de données</b>	 <b>94</b>
 <b>D Incertitude statistique et incertitude de mesure</b>	 <b>99</b>
 <b>E Article soumis à PRE</b>	 <b>101</b>
 <b>Bibliographie</b>	 <b>121</b>



## Première partie

### Introduction : l'expression génétique chez *Escherischia Coli*

La bactérie *Escherichia Coli* (*E. Coli*) est un être vivant unicellulaire procaryote d'environ quelques  $\mu\text{m}^3$ . Elle a été découverte dans des selles infantiles au 19<sup>ème</sup> siècle par Theodor Escherich et forme la flore intestinale de pratiquement tous les êtres vivants dotés d'appareil digestif. Certaines souches d'*E. Coli* sont responsables de pathologies telles que les gastro-entérites, les méningites ou encore les septicémies. Par la suite, *E. Coli* est devenue un formidable outil d'investigation pour la génétique moléculaire de par sa rapidité de reproduction et la facilité avec laquelle on peut modifier ses propriétés génétiques. Ceci est d'autant plus vrai que son génome a été complètement séquencé.

égulation de la rplication

# Chapitre 1

## Le dogme central

Dans le dogme central de la biologie moléculaire [1], l'acide désoxyribonucléique (ADN) est le support de l'information génétique. L'ADN est un polymère dont chacun des monomères est constitué d'un sucre phosphaté et d'une base azotée : Adénine (A), Cytosine (C), Guanine (G) ou Thymine (T). Les monomères d'ADN peuvent être liés de manière covalente au niveau des sucres afin de former un simple brin d'ADN orienté<sup>1</sup>. De plus, les bases peuvent s'apparier deux à deux en formant plusieurs liaisons hydrogènes<sup>2</sup> : A avec T et C avec G. Si deux brins sont complémentaires l'un de l'autre, ils peuvent former une double hélice d'ADN.

Ainsi, l'ADN, de par la séquence de ses bases azotées, code non seulement pour sa propre réplication, mais contient aussi toute l'information nécessaire au fonctionnement de la cellule.

Tout d'abord l'ADN est transcrit en acide ribonucléique ARN. Cette étape est la transcription. Cet ARN peut soit être utilisé comme catalyseur pour une réaction biochimique soit être dégradé, soit être utilisé pour réguler certains mécanismes d'expression, ou encore être la matrice à partir de laquelle on fabriquera les protéines. Le passage d'information de l'ARN vers la protéine est la traduction.

Deux flux d'informations sont prépondérants : de l'ADN vers l'ARN puis de l'ARN

---

<sup>1</sup>Les simples brins d'ADN sont orientés car la liaison entre deux bases d'ADN se fait par l'intermédiaire d'un phosphate entre le carbone 3 du premier cycle et le carbone 5 du deuxième cycle. On parle alors d'orientation 5' vers 3' pour une chaîne d'ADN simple brin car sa première extrémité présente un phosphate sur le carbone 5 de son cycle et un hydroxyle à l'autre extrémité sur le carbone 3.

<sup>2</sup>L'énergie des liaisons hydrogène entre deux bases est de 2 kT pour une paire A-T et de 3 kT pour une paire G-C.

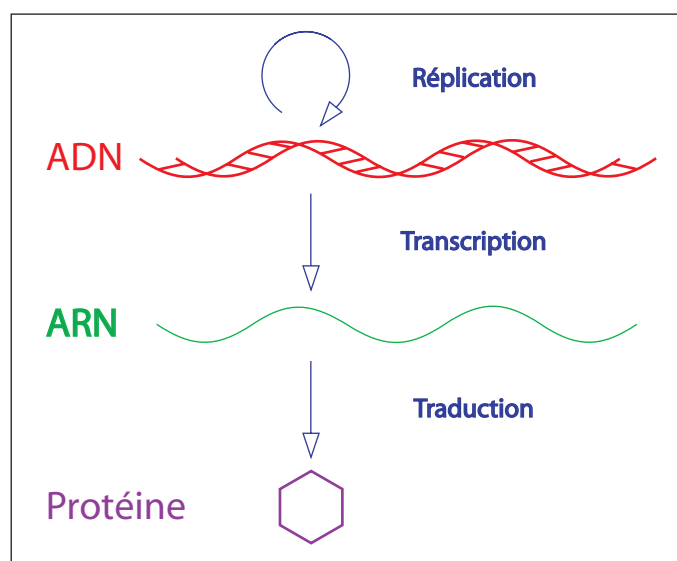


FIG. 1.1: Le dogme central s'articule autour de trois processus : la réplication, la transcription et la traduction.

vers les protéines. Cependant, un certain nombre de phénomènes viennent souligner que le transfert d'information peut se faire autrement. L'un des contre exemples les plus flagrants est l'existence des rétrovirus. Ce sont des virus dont le génome est fait d'ARN. Lorsque ces rétrovirus infectent un organisme, ils «convertissent» leur ARN en ADN et insèrent l'ADN au sein du chromosome de l'hôte.

Même si ce contre-exemple est loin d'être le seul, le dogme central est un cadre assez concis qui permet d'aborder l'expression génétique. Les flux d'information les plus importants se font donc par la transcription et la traduction que nous allons détailler.

## 1.1 La transcription

Lors de la transcription, un brin d'ARN messenger (ARNm) est produit partir d'une matrice d'ADN.

L'ARN polymérase (ARNpol) reconnaît une séquence spécifique d'ADN, le promoteur. Plus précisément, il reconnaît principalement au sein du promoteur la «Pribnow box» (*TATAAT*). L'ARNpol ouvre alors la double hélice et commence la transcription.

L'ARNpol lit une base du brin transcrit de l'ADN dans le sens 5' vers 3', lui accole la base ARN complémentaire, et lie cette dernière à la base ARN rajoutée précédemment.

Elle avance d'un cran et recommence.

La transcription s'arrête lorsque la polymérase arrive à un site de terminaison. Ce dernier lorsqu'il est transcrit provoque le décrochage de la polymérase par le biais de l'ARNm qui forme une boucle en épingle à cheveux qui déstabilise le complexe ADN-ARNPol [2].

Le taux de transcription d'un gène est contrôlé par le promoteur qui lui est adjoint. Non seulement, celui-ci doit contenir le site d'accrochage de la polymérase, mais il contient généralement d'autres séquences régulatrices. Ces séquences peuvent modifier les conditions dans lesquelles la transcription est possible ou moduler l'intensité de la transcription en fonction de l'environnement<sup>3</sup>. Nous en verrons un exemple concret plus loin. Retenons simplement que la régulation d'un gène peut se faire via la régulation dite transcriptionnelle et que cela constitue une grande partie de la régulation procaryote

## 1.2 La traduction

Une fois l'ARNm synthétisé, un ribosome s'attache à l'ARNm au niveau d'une séquence qu'on appelle le «ribosome binding site» (RBS) ou «Shine-Dalgarno» (*AGGAGG*).

Ensuite, il reconnaît le triplet (codon) *AUG*<sup>4</sup> à  $\sim 8$  bases en aval et commence à traduire la séquence d'ARNm 3 nucléotides par 3 à partir de celui-ci. Chaque codon est associé à un ARN de transfert (ARNt). Chaque ARNt contient une zone de reconnaissance (anti-codon) correspondant à un seul codon de l'ARNm et l'acide aminé correspondant au codon reconnu. L'ARNt et son acide aminé ne sont utilisés par le ribosome que si l'anticodon de l'ARNt correspond au codon lu. L'acide aminé porté est alors fixé de manière covalente à l'acide aminé du codon précédemment lu. Le ribosome se déplace ensuite de trois bases et lit le codon suivant.

La traduction s'arrête lorsque le ribosome rencontre l'un des codons stop ou «non-sens» (*UAA*, *UAG* ou *UGA*) pour lesquels aucun ARNt correspondant n'existe. A partir de là, la chaîne peptidique est relarguée dans le cytoplasme. Son repliement tridimensionnel et

---

<sup>3</sup>Par environnement, nous désignons à la fois les séquences d'ADN présentes autour du gène considéré et les composés présents localement dans le cytoplasme. Le milieu extérieur de la bactérie peut bien sûr influencer le contenu du cytoplasme dans la limite des composés auxquels la membrane est perméable qu'il s'agisse de transport passif ou actif et de leurs dérivés.

<sup>4</sup>Ce triplet est souvent appelé codon «start»

sa maturation terminent de rendre la protéine fonctionnelle.

### 1.3 L'opéron Lactose, un exemple de régulation génétique.

Les gènes forment des réseaux d'interactions et régulent leur expression entre eux. Afin d'illustrer comment des interactions peuvent se mettre en place, nous prendrons l'exemple, désormais très bien étudié, de l'opéron Lactose<sup>5</sup>. Un opéron est un ensemble de gènes œuvrant pour une même fonction.

De plus, l'opéron Lactose est le premier groupe de gènes qui montrait explicitement une régulation de certains gènes fonction des conditions environnementales à avoir été mis en évidence. Il permet d'utiliser le lactose comme source de carbone. L'opéron comporte une région régulatrice, composée du gène *lacI* et d'un promoteur, suivie de trois gènes structuraux(*lacZ*, *lacY* et *lacA*)<sup>6</sup>.

**L'effet du lactose** Le gène *LacI* est exprimé de manière constitutive<sup>7</sup>. Il donne lieu la formation d'un tétramère de protéines répresseur. Ce dernier peut alors former un complexe avec un site spécifique de l'opéron ce qui empêchera la transcription des gènes structuraux.

En présence de lactose, la répression de l'opéron est arrêtée. En effet, une forme isomérique du lactose, l'allolactose<sup>8</sup>, peut se lier chacun des monomères du répresseur ce qui changera la conformation du tétramère, l'empêchant ainsi de se lier au site spécifique de l'opéron. Ceci a pour conséquence la levée de l'inhibition des gènes *lacZ*, *lacY*, et *lacA*.

---

<sup>5</sup>LA découverte de l'opéron lactose par Jacob et Monod est la première preuve d'une régulation de l'expression des gènes par le produit d'autres gènes. Quelques animations permettant une visualisation directe du fonctionnement de l'opéron peuvent être trouvées sur le site suivant : <http://www.snv.jussieu.fr/bmedia/operonlactose/>

<sup>6</sup>Les gènes *lacZ*, *lacY* et *lacA* codent pour les protéines nécessaires à l'utilisation du lactose comme source de carbone. La  $\beta$ -galactosidase (gène *lacZ*) et la lactose perméase (gène *lacY*) permettent respectivement la transformation du lactose en glucose et l'entre massive du lactose au sein de la cellule. Quant à la thiogalactoside transacétylase (gène *lacA*), son rôle reste mal défini.

<sup>7</sup>la production n'est pas régulé. Une expression constitutive désigne une expression permanente. Cependant, le niveau d'expression n'est pas déterminé de manière précise.

<sup>8</sup>La  $\beta$ -galactosidase catalyse l'isomérisation de lactose en allolactose.

**L'effet du glucose** Cependant, la présence de lactose dans le milieu n'est pas suffisante pour la production de LacY, LacZ et LacA. Même si en présence de lactose, les gènes structuraux peuvent être transcrits, leur taux de transcription est très faible. Afin d'obtenir une transcription non négligeable, un complexe de protéines est nécessaire : il s'agit de la protéine CAP et de l'AMPc qui est produit lorsque le glucose se raréfie<sup>9</sup>. Ce complexe favorise le recrutement de l'ARN polymérase car sa fixation au promoteur augmente l'affinité de l'ARN polymérase pour le promoteur. Une forte transcription est donc conditionnée par le manque de glucose.

L'opéron lac permet la synthèse des enzymes nécessaires seulement si le lactose est présent et si le glucose manque. Son étude met en évidence une première voie de régulation par inhibition et une deuxième voie par activation. L'inhibition se fait par fixation d'une protéine un site, ce qui empêche la fixation d'autres acteurs sur ce même site voir sur les sites environnants. L'inhibition empêche alors l'appareillage de la transcription d'accomplir la tâche qui lui est impartie. L'activation, elle, se fait sur le même mode de fonctionnement. La protéine activatrice se fixe sur un site, mais de par son affinité avec une protéine effectrice, elle la recrutera afin de l'approcher de sa cible.

La relative simplicité de fonctionnement de l'opéron permet d'entrevoir les mécanismes dont dispose la bactérie afin de moduler son activité en fonction de son environnement. De plus, si dans nos explications, les signaux d'entrée de l'opéron étaient des signaux "extérieurs", dans un cas plus général, les signaux des différents sous réseaux d'expression génétique peuvent provenir d'autres sous réseaux. Ainsi, des comportements assez riches peuvent apparaître<sup>10</sup>.

Dans notre cas, nous nous intéresserons à l'expression de manière quantitative, au sein de populations isogéniques<sup>11</sup> de bactéries. Ainsi, nous nous concentrerons sur les variations d'expression de manière quantitative, et sur certains réseaux de régulation.

---

<sup>9</sup>L'AMPc est un signal biochimique de famine.

<sup>10</sup>Pour une lecture plus complète de l'opéron lactose, nous invitons le lecteur parcourir la revue de Vilar et al.[3]

<sup>11</sup>Une population isogénique est une population où l'ensemble des individus ont le même génome.

## Chapitre 2

# Les variations phénotypiques et les réseaux

Au sein d'une population isogénique, plusieurs phénomènes peuvent être source d'individualité. Si on met de côté le cas où la population n'est pas exposée aux mêmes conditions, la première source à laquelle nous pensons est le hasard des mutations génétiques qui altèrent l'information que porte chacune des bactéries.

**Les variations génétiques** qui émergent dans une population isogénique peuvent être estimées grâce aux nombreuses données sur le sujet (trouver une citation). Ainsi, le taux de mutation est  $\sim 10^{-10}$  par paire de bases par génération [4]. Pour un gène d'environ un millier de paires de bases, on obtient au final un taux de mutation au sein d'un gène de  $\sim 10^{-7}$  par génération. Supposons que toutes les mutations sont significatives. On peut trouver  $\sim 10^8$  bactéries dans un millilitre de milieu nutritif saturé en bactéries. Ce qui implique que sur la population qui nous intéresse, on trouvera une ou deux dizaines d'individus par ml dont le gène aura été modifié. Plus précisément, si on mesure l'expression d'un gène spécifique sur  $\sim 10^7$  bactéries, seule une ou deux d'entre elles auront subi une mutation sur le gène en question. Ce phénomène est le moteur de l'évolution. Dans les expériences d'évolution de Bennett et Lenski [5], les temps nécessaires sont de l'ordre du millier de générations, Cependant, aux petites échelles de temps (quelques dizaines de générations), il n'est pas nécessaire d'en tenir explicitement compte<sup>1</sup>.

---

<sup>1</sup>Il peut être nécessaire d'éviter les dérives génétiques ou l'accumulation de mutations. Pour cela, il suffirait simplement de partir du même clone systématiquement.



Par contre, les variations qui vont nous intéresser sont celles qui émergent malgré l'uniformité des conditions externes et du matériel génétique. Ainsi, dès 1976, Spudich et Koshland [6] montraient que les fluctuations stochastiques dans les réactions biochimiques pouvaient donner naissance à une certaine individualité. Le petit nombre de molécules d'une espèce biochimique dans une bactérie peut engendrer des fluctuations importantes de quantité que l'on ne pourra négliger. Plusieurs travaux ont depuis montré l'importance de la stochasticité dans l'expression génétique. ([7], [8], [9]).

Nous nous attacherons dans ce chapitre à expliciter quelques caractéristiques de ces variations phénotypiques et l'influence que la structure du réseau d'interactions peut avoir sur la variabilité. Nous finirons alors par une brève discussion sur les liens entre variabilité et adaptabilité.

## 2.1 Variation d'expression d'un gène

Chacune des réactions biochimiques impliquées dans l'expression d'un gène s'effectue lorsque les composants de cette réaction soumis à l'agitation thermique sont "suffisamment proches" les uns des autres. Ce caractère aléatoire impose une stochasticité dans les réactions dont on doit tenir compte.

L'expression d'un gène est alors soumise à certaines fluctuations de manière intrinsèque. Afin d'explicitier quelques caractéristiques de cette variabilité, prenons le modèle d'expression développé dans Thattai et van Oudenaarden [10] (voir figure 2.1). Le gène considéré est transcrit de manière constitutive avec un taux  $k_r$ . L'ARN produit est alors traduit avec un taux  $k_p$ . Les ARN et protéines sont dégradés respectivement avec les taux  $\gamma_r$  et  $\gamma_p$ . Dans

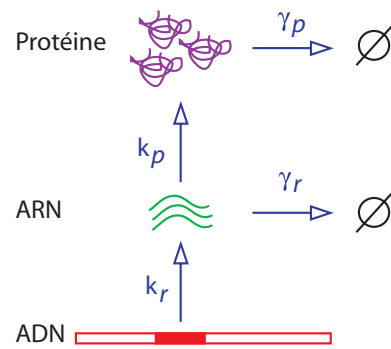


FIG. 2.1: Modèle d'expression génétique. avec des taux de réaction constants. Reproduit à partir de [10]

ce modèle, tous les paramètres sont constants et les autres phénomènes sont négligés. Ainsi, on ne tient pas compte de la disponibilité du promoteur, du nombre de polymérases ou de leurs variabilités. Chaque réaction est alors considéré comme un processus de

poisson. Le système est alors décrit à l'aide de probabilité d'être dans un certain état et des probabilités de transition d'un état vers un autre.

Les auteurs déduisent de ce modèle le niveau moyen d'expression de la protéine,  $\langle p \rangle$ , et la variabilité de son expression qu'ils quantifient à l'aide du facteur de Fano,  $\frac{\delta p^2}{\langle p \rangle}$ , à l'état stationnaire :

$$\langle p \rangle = \frac{k_r \cdot b}{\gamma_p} \quad (2.1)$$

$$\frac{\delta p^2}{\langle p \rangle} = 1 + b \quad (2.2)$$

où  $b = \frac{k_p}{\gamma_r}$  exprime le nombre de protéines produites à partir du même ARN.

On peut calculer à partir de ces formulations analytiques la valeur de la production de protéine et sa variance pour des paramètres donnés à l'état stationnaire. La configuration de base tiré de [10] utilise les paramètres suivants :  $k_r = 0.01s^{-1}$ ,  $\gamma_r = \ln(2)/120s^{-1}$ ,  $k_p = 20 \cdot \gamma_r$ ,  $\gamma_p = \ln(2)/3600s^{-1}$ . En fait prendre de telles valeurs revient à poser qu'un ARNm est produit toutes les 100 secondes, que le temps de vie d'un ARNm est de  $\sim 2$  minutes, que pour chaque ARNm, on produit environ 20 protéines et que chaque protéine a un temps de vie typique de l'ordre de l'heure. On trouve alors  $\langle p \rangle \simeq 1038$  et  $\frac{\delta p^2}{\langle p \rangle} \simeq 21$ . On obtient alors une variabilité ou un bruit intrinsèque<sup>2</sup>  $\eta_i = \sqrt{\frac{\delta p^2}{\langle p \rangle^2}} \simeq 14\%$ . La figure 2.2 donne une idée graphique de l'étendue du bruit d'un tel schéma d'expression.

Dans le cas d'une expression constitutive non régulée, nous avons donc une variabilité importante. Cependant, il est possible au sein d'un réseau de gènes de moduler cette variabilité.

Les gènes font partie de réseaux d'interactions au sein desquels ils se contrôlent mutuellement et conditionnent le métabolisme de la cellule. On peut alors se demander à quel point leur expression fluctue, si l'expression d'un gène fluctue autant que celle d'un autre, si une réaction en chaîne amplifie ou minimise la variabilité, ou encore quels sont les mécanismes par lesquels l'organisme minimise ses variations.

---

<sup>2</sup>«variabilité intrinsèque» désigne la variabilité **directement** liée à l'expression du gène dont on tient compte.

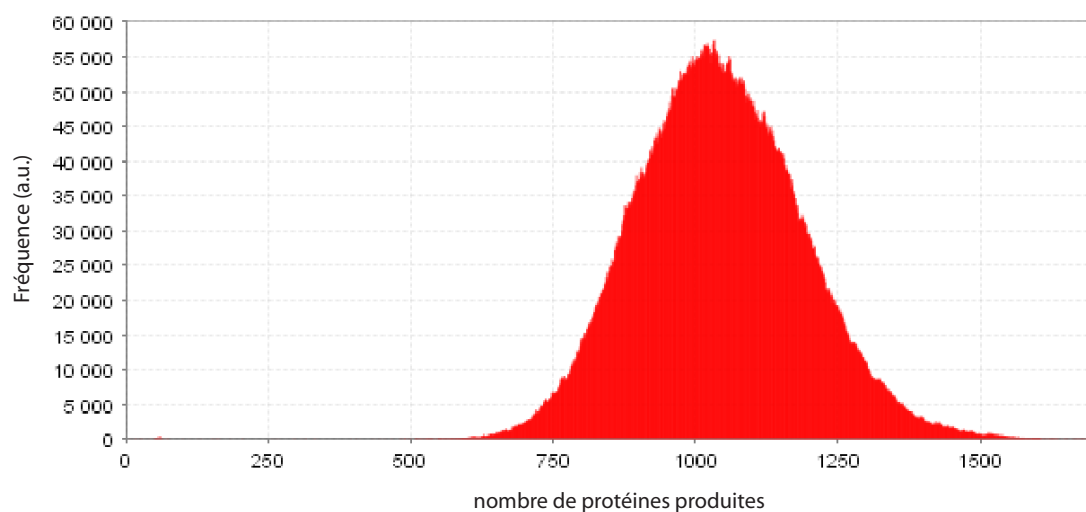


FIG. 2.2: Histogramme du nombre de protéines exprimées selon le schéma d'expression de la figure 2.1. Cet histogramme a été généré avec un programme de simulation stochastique, BioNetS 2.0 [11].

## 2.2 Variabilité conditionnée de l'expression génétique.

La démonstration la plus simple d'une variabilité dépendante de la structure du réseau est apportée par les travaux de Elowitz et al. [9]. Deux gènes codant pour deux protéines fluorescentes (la *Yellow Fluorecent Protein* (yfp) et la *Cyan Fluorecent Protein* (cfp)) ont été clonés<sup>3</sup> dans le chromosome d'une bactérie. Ces deux gènes étaient sous le contrôle de promoteurs identiques. La mesure des intensités de fluorescence dans le spectre d'émission de la yfp ou de la cfp permettait d'extraire les variations intrinsèques à l'expression du gène d'intérêt.

Les auteurs font une distinction entre deux sources de "bruit". Lorsque l'on considère les variations entre les bactéries d'une même population isogénique, il est possible de disséquer la variabilité totale  $\eta_{tot}$  en variations intrinsèque  $\eta_i$  et extrinsèque  $\eta_e$ . La première est directement issue de la stochasticité de l'expression du gène que l'on considère. Ainsi, selon que la production soit constitutive, réprimée ou induite, les auteurs trouvaient des variations intrinsèques  $\eta_i$  allant de 5 à 25%. La deuxième quant à elle provient de l'accumulation des différences entre les différentes cellules de la population (nombre de polymérases, de ribosomes, etc.). Elowitz et al. trouvent ainsi une variabilité extrinsèque

---

<sup>3</sup>cloné ici signifie inséré

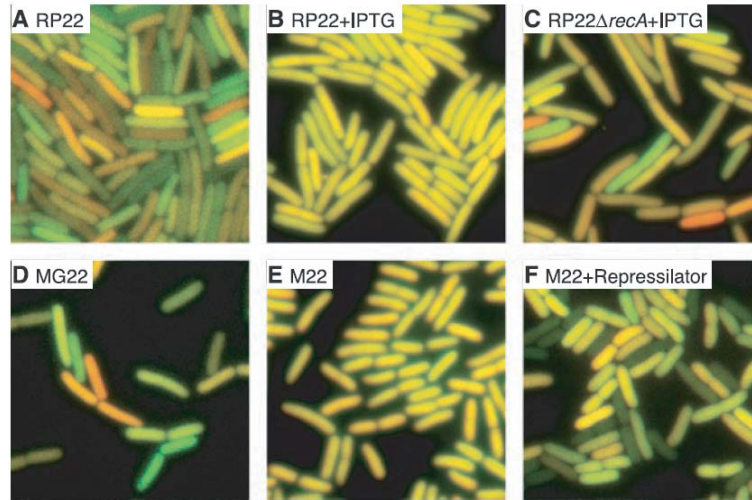


FIG. 2.3: L'expression de la *cfp* et celle de la *yfp* sont combinées sur une même image en vert et en rouge. Ainsi, une expression équivalente donne une couleur jaune. Les différentes images sont différentes situations biologiques de l'expression génétique. On trouve ainsi une population assez uniforme pour les images B et E, alors qu'on retrouve une grande hétérogénéité pour toutes les autres configurations. Ces images ont été reprises de Elowitz et al. [9]

$\eta_e$  allant de 2,4 à 57%. Il s'agit en fait d'une extension du travail de Thattai et al.[10] qui donnaient une définition de la variabilité intrinsèque.

Lorsque les gènes rapporteurs sont sous le contrôle du promoteur Lac, pouvant être réprimé par la protéine *lacI*, une forte variabilité est constatée (Figure 2.3 A). Lorsqu'on lève l'inhibition en rajoutant de l'isopropyl-thiogalactosidase (IPTG) ou en supprimant le gène *lacI*, le bruit s'en trouve réduit d'un facteur 5 (Figure 2.3 B et E). De plus, la délétion d'un gène particulier mène à une augmentation de la variabilité (Figure 2.3 C).

Il fut donc apporté la preuve expérimentale qu'en fonction du promoteur, des régulateurs de ce promoteur, et de l'environnement<sup>4</sup>, la variabilité d'expression au niveau de la population  $\eta_{tot}$  peut passer de 8 à 58%.

Il est donc nécessaire lorsqu'on veut mesurer l'activité d'un gène de tenir compte du réseau d'interactions qui régule son expression. Ainsi, Thattai et van Oudenaarden [10]

<sup>4</sup>On désigne ici aussi bien l'environnement dans lequel est plongé la population de bactérie que l'environnement de l'expression au sein de la bactérie (le cytoplasme) avec toutes les protéines qui peuvent interagir avec l'expression considéré.

montrèrent que différentes structures régulatrices menaient à des variabilités différentes. Au sein de leur modèle, plusieurs structures furent ainsi comparées :

- l'expression non régulée servait d'étalon.
- la boucle d'autorégulation négative donnait une variabilité moindre.
- la boucle d'autorégulation positive offrait deux états stables<sup>5</sup>.
- une paire de gène se réprimant mutuellement faisait apparaître de la bistabilité, avec une probabilité non nulle de passer de l'un vers l'autre.
- la cascade de gènes inducteurs propageait et amplifiait le bruit.

On peut alors se demander quel rôle joue le bruit au sein des réseaux de gènes.

## 2.3 Variabilité : bruit ou adaptabilité

L'expression génétique fait intrinsèquement apparaître de la variabilité. Une fonction génétique particulière peut donc être extrêmement variable, "bruitée". Par contre, certains réseaux semblent pouvoir s'affranchir du bruit d'expression ou d'autres perturbations. On peut alors se demander s'il existe des cas où ce bruit peut être avantageux au niveau de la population et si certaines structures de réseau ont été choisies afin d'accentuer ou de diminuer la variabilité.

### 2.3.1 Fonction et robustesse

Un grand nombre de fonctions d'*E. Coli* sont contrôlés par des réseaux génétiques. Il en va ainsi de l'utilisation du lactose comme source de carbone. Cette fonction est conditionnée par l'opéron lactose. Cependant, il est nécessaire afin de conserver la fonction de maintenir le réseau (exprimer le gène LacI de manière constitutive par exemple), ce qui implique un coût métabolique.

Les réseaux de ce type délivrent généralement une réponse adéquate malgré la variabilité de certaines grandeurs (et ce sur plusieurs ordres de grandeurs quelques fois) : Ils sont robustes. La mesure effectuée afin de caractériser le milieu environnant et la réponse du réseau à des changements dans le milieu sont alors extrêmement fiables. Lors du développement d'organismes multicellulaires, la concentration de certaines protéines déterminent

---

<sup>5</sup>Leur calcul reste assez obscure pour moi... il faudrait vraiment refaire proprement leur cheminement afin de comprendre comment ils arrivent à ce résultat.

le devenir de certaines cellules. Le déterminisme strict de la différenciation cellulaire de certains organismes tels que *C. Elegans* montre la robustesse de tels réseaux.

Pour de tels réseaux, la variabilité d'expression apparaît comme un bruit diminuant leur performance ou pouvant induire des erreurs. Il est alors intéressant d'étudier la structure de ces réseaux afin de mettre en évidence comment le réseau peut rester robuste face au bruit d'expression génétique.

Le travail exposé dans Shenn-Orr et al. [12] montre la présence de motifs au sein du réseau d'interactions transcriptionnelles d'E.Coli. Il est alors intéressant de noter que 27 des 64 facteurs de transcription soit plus de 40% exhibent une autorégulation négative ce qui nous l'avons vu plus haut diminue la variabilité d'expression de ces mêmes facteurs de transcription.

Cependant, il existe des cas où le rôle de la variabilité peut s'avérer plus ambigu.

### 2.3.2 Choisir la variabilité plutôt qu'un contrôle global

On peut imaginer que dans certains cas, il soit avantageux de créer de la diversité. Prenons l'exemple de la chemotaxie. Si toute la population répond de la même manière un attracteur chimique, elle remontera en bloc son gradient. Comme suggéré par Spudich et Koshland [6], il se peut alors que toute la colonie se retrouve dans un milieu toxique qu'elle ne sait pas détecter ou qu'elle soit exposée à un prédateur. Ainsi, l'individualité de la réponse à l'attracteur peut apparaître comme une véritable stratégie de survie, le «bet-hedging».

Cette stratégie a déjà été mise en évidence dans des systèmes écologiques faisant intervenir des comportement animaliers [13]. De plus certaines situations dans lesquels une diversité phénotypique apparaît au sein de populations isogéniques sont souvent interprétés en ces termes. On pense alors au cas du «switch» de la persistance bactérienne [14] : La bactérie peut passer dans un état où elle croît moins vite mais dans lequel elle est résistante à certaines doses d'antibiotiques. Au sein de la population, seule une toute petite partie d'entre elles passent dans cet état. De manière globale, les bactéries sont adaptées au milieu sans antibiotique. Par contre, lorsqu'il arrive un flux d'antibiotique temporairement, toute la colonie meure à l'exception des bactéries persistantes qui, de manière probabiliste, peuvent redonner des bactéries non persistantes.

Le bet-hedging ou la stratégie de diversification est souvent associé à des changements

de milieu. En effet, lorsque le milieu est fluctuant (rapidement) ou face à un changement environnemental important et non prévisible, on peut imaginer que générer de l'hétérogénéité de manière stochastique permette à une fraction de la population de survivre. Ainsi, Kussel et al. [15] [16] ont montré que dans certains cas (notamment par le choix des paramètres utilisés), le choix de générer de la diversité par des fluctuations était avantageux par rapport à une voie de signalisation extrêmement contrôlée qui permettrait une adaptation de l'ensemble de la colonie au changement. En effet, un «switch» stochastique sautant dans un état adapté ou non au milieu permettait à la population d'avoir un taux de croissance plus important qu'une voie de signalisation qui mesurerait l'état du milieu et qui adapterait l'état de chaque bactérie à son environnement.

La variabilité peut donc être un facteur d'adaptabilité au niveau de la population.

## Chapitre 3

# Le chromosome d'*E. Coli*

Toutes les cellules vivantes contiennent au moins un chromosome codant pour l'information nécessaire à son métabolisme. Le chromosome d'*E. Coli* est une double hélice bouclée sur elle-même, formée de deux simple brins/polymères d'ADN. Le chromosome d'*E. Coli* est ainsi constitué de 4,6 millions de paires de bases (bp).

La séquence des différentes bases d'ADN est l'information portée par le chromosome. Il est possible de distinguer au sein du chromosome des ensembles de bases, des gènes. Ces gènes codent pour l'information nécessaire à la production des protéines qui présentent généralement des interactions spécifiques avec les autres espèces biochimiques présentes. De plus, chez la bactérie, la plupart des gènes sont exprimés au sein d'«opérons», groupes de gènes opérant ensemble pour une fonction précise de l'organisme. Nous avons d'ailleurs déjà vu l'exemple de l'opéron lactose.

La bactérie *E. Coli* ne contient qu'un seul exemplaire de son chromosome. Afin de se reproduire, la bactérie doit donc dupliquer son chromosome. Ensuite, elle se divise pour donner naissance à deux bactéries filles contenant *a priori* le même matériel génétique.

Nous détaillerons dans ce chapitre deux aspects de la reproduction bactérienne : La duplication du chromosome et son partitionnement entre les deux bactéries filles.

### 3.1 La réplication du chromosome

La réplication du chromosome est un phénomène biochimique faisant intervenir un grand nombre d'espèces et variant selon l'organisme auquel on s'attache. Dans le cas de la bactérie *E. Coli*, c'est l'ADN polymérase III qui assure cette fonction à l'aide d'un



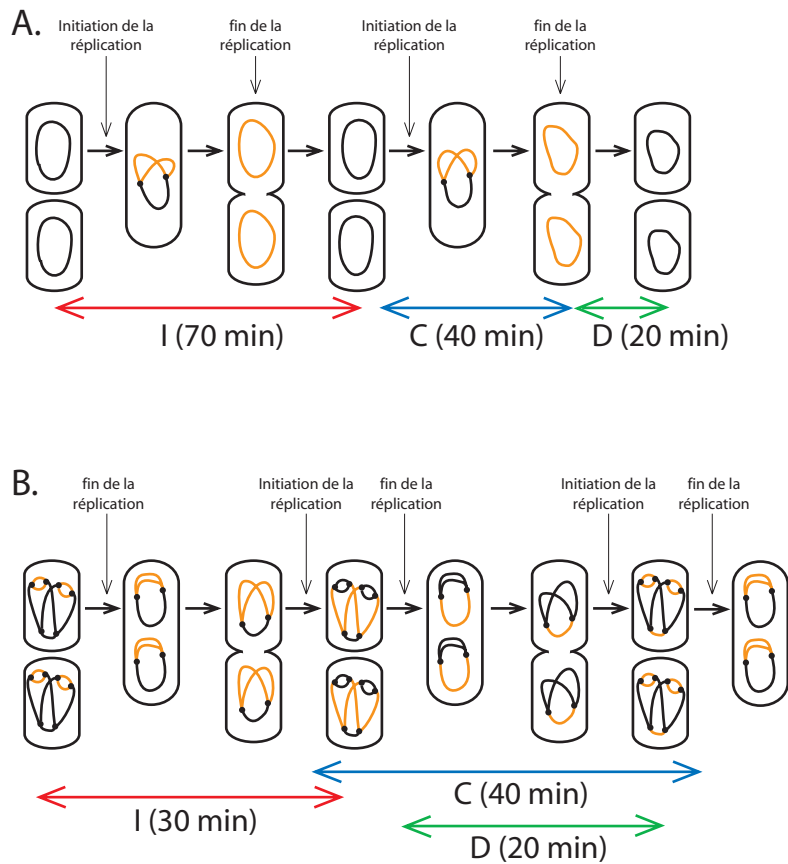


FIG. 3.1: Cycle de reproduction de la bactérie. Reproduit à partir de [17]. La duplication du chromosome est nécessaire pour la reproduction de la bactérie. I, C, et D dénotent respectivement le temps entre deux divisions, le temps nécessaire la réplication et le temps qui s'écoule entre la fin de la réplication et la répartition des deux copies du chromosome dans chacune des bactéries filles. (A)  $C+D$  est inférieur à I. Le chromosome de la bactérie est dupliqué, ensuite la division a lieu. (B)  $C+D$  est supérieur à I. Cette situation arrive notamment en milieu riche. Il n'y a toujours qu'une seule initiation et une seule terminaison de la réplication par cycle. Par contre, la terminaison qui a lieu durant un cycle a été initiée lors du cycle précédent.

certain nombre d'autres protéines tel que l'hélicase DnaB<sup>1</sup> ou encore la primase DnaG<sup>2</sup> ([17],[18]).

### 3.1.1 La régulation de la réplication

L'initiation de la réplication se fait à un endroit précis (séquence déterminée) du chromosome, l'*OriC*. L'initiation doit aussi avoir lieu une fois par *OriC* par cycle cellulaire. De plus, ce processus est extrêmement long et coûteux à la bactérie. Il est donc nécessaire à la bactérie de réguler ce processus afin qu'il n'induisse pas de coût superficiel.

Parmi toutes les protéines impliquées dans la réplication, DnaA est au cœur du processus. Sa concentration semble jouer un rôle prépondérant dans la régulation de la réplication [19]. Elle se fixe à des sites spécifiques. Ensuite, les DnaA (10 à 12) s'agrègent et forment un complexe avec l'ADN qui déstabilise la forme double brin adjacente au profit d'une forme ouverte présentant deux simples brins. D'autres protéines formant la fourche de réplication peuvent alors s'associer à la forme ouverte de l'ADN, ce qui achève d'initier la réplication. Au début du cycle cellulaire et ce pendant une certaine période, l'initiation n'est pas possible étant donné que DnaA n'a pas encore été suffisamment produite. Une fois, le seuil «d'activation» dépassé<sup>3</sup>, l'initiation de la réplication a lieu.

Pour le reste du cycle cellulaire, afin d'éviter une deuxième initiation sur la même origine de réplication, sont principalement mis à contribution deux mécanismes : Le premier est en fait l'auto-régulation négative de DnaA [20] (Si DnaA est trop produite, son expression s'arrête). Le deuxième consiste en un appareillage de séquestration empêchant l'accès de certaines protéines à l'*OriC*. En effet, le brin d'ADN initial est méthylée<sup>4</sup> alors que les brins naissant ne le sont pas encore. La protéine SeqA reconnaît les double-brins hémi-méthyles donc nouvellement produits, les séquestre à la paroi et empêche d'autres complexes protéiques d'avoir accès à l'origine de réplication [21].

---

<sup>1</sup>L'hélicase DnaB sépare les deux brins formant la double hélice d'ADN

<sup>2</sup>La primase DnaG synthétise les amorces ARN sur un des brins d'ADN

<sup>3</sup>Le seuil d'activation correspond à un seuil de concentration de DnaA. Au début du cycle cellulaire, DnaA est produite continuellement, avec le même taux. Sa concentration varie linéairement. Le seuil d'activation correspond alors à un certain délai de temps, l'«eclipse time».

<sup>4</sup>La méthylation consiste à rajouter un groupement méthyle sur un des carbones du squelette de l'ADN. On change alors l'accessibilité de l'ADN aux protéines. De plus, la réplication se fait de manière semi-conservative. Chacun des brins est complété afin d'obtenir deux double-brins. Comme le brin nouvellement synthétisé n'est pas encore méthyle, nous obtenons donc deux double brins hémi-méthylés.

### 3.1.2 La fourche de réplication

A la suite de l'initiation de la réplication, deux fourches de réplication sont formées à l'*OriC*. Elles vont ensuite progresser chacune en sens opposé. La fourche de réplication ouvre le double brin d'ADN et complète chacun des simple brins ; au fur et à mesure que les fourches progressent le long du chromosome, une plus grande partie de celui-ci est dupliquée. La réplication se termine lorsque les deux fourches collisionnent.

Les fourches de réplication progressent à  $\sim 1000\text{bp/s}$  : la duplication du chromosome prend donc  $\sim 40$  minutes. Pour des temps de division plus grands que le temps nécessaire à la réplication, deux copies du chromosome sont présentes en fin de cycle au sein de la bactérie. Par contre, en milieu riche à  $37^\circ\text{C}$ , le temps de division d'*E. Coli* est de  $\sim 20$  minutes. A ce moment là, même si le nombre de fourches de réplication en cours triple ou plus, la fréquence d'initiation, elle, reste la même. En effet, une duplication, qui commence en début de cycle d'une bactérie, ne se termine qu'à la fin du cycle de ses bactéries filles. Il aura donc été nécessaire au début du cycle que la bactérie ait non pas deux fourches de réplication mais six dont deux ayant déjà répliqué la moitié du chromosome .

La caractéristique la plus marquante de la réplication est son extrême fidélité. En effet, la polymérase agit avec un taux d'erreur de  $10^{-7}$  par paire de bases<sup>5</sup> [4] , soit moins une erreur en moyenne sur l'ensemble du chromosome par réplication. Elle garantit ainsi que la même information se conserve de génération en génération.

Afin que la scissiparité soit effective, il faut que chacune des bactéries filles contienne une copie du chromosome.

## 3.2 La partition du chromosome

Plusieurs gènes chez *Bacillus Subtilis* permettent une répartition correcte des deux copies du chromosome entre les bactéries filles. Cependant, chez *E. Coli*, aucun de ces gènes ou leurs homologues n'a pu être identifié.

Récemment, des simulations numériques ont montré que deux polymères fermés peuvent ségréger<sup>6</sup> s'il sont suffisamment contraints dans une direction. De plus, la réplication pro-

---

<sup>5</sup>Pour obtenir ce taux d'erreur, on tient compte de la capacité de vérification de la polymérase de *E. Coli* (polIII)

<sup>6</sup>Les deux polymères n'occupent pas le même espace physique

gressive du chromosome/polymère permettrait de rendre compte de la dynamique de positionnement du chromosome si l'on tient compte du positionnement des fourches de réplication [22].

Bien que la répartition se fasse «par elle-même», d'autres mécanismes empêchent que la zone de constriction, le septum, ne se referme en coupant une partie du chromosome. Ainsi, une batterie de protéines incluant le système mukFEB ou encore la protéine FtsZ permettent d'éviter ces accidents au chromosome[23].

## Chapitre 4

# Les plasmides

Un plasmide est un fragment circulaire d'ADN extra-chromosomal n'exprimant *a priori* pas de fonctions essentielles à la bactérie. Cependant, les plasmides peuvent conférer à leur hôte des aptitudes qui dans certains cas lui donnent un avantage sélectif. Cet avantage sélectif peut provenir d'une résistance à un antibiotique ou encore de la possibilité d'utiliser une nouvelle voie métabolique.

Il existe ainsi une grande variété de plasmides. Ils diffèrent non seulement par la multiplicité des apports à leur hôte, mais aussi par les mécanismes qui les maintiennent dans la population. La principale caractéristique à laquelle nous nous intéresserons est leur capacité à contrôler leur nombre de copies au sein de la bactérie ou de la population. Les mécanismes qui sont liés à la réplication du plasmide et ceux qui permettent une répartition équilibrée des copies entre les deux bactéries filles lors de la division cellulaire seront les éléments dominants de notre étude.

Nous avons fait le choix de certains plasmides qui nous apparaissaient comme bien décrits du point de vue biologique [24]. Je détaillerais tout d'abord la régulation du nombre de copies de plasmides au sein de chacun des plasmides choisis. Je terminerais ensuite par une brève discussion sur la variabilité du nombre de copies de plasmides.

### 4.1 Le Plasmide F

Le F est un plasmide de taille élevée ( $\sim 100\text{kbp}$ ) qui a joué un rôle très important aux débuts de la biologie moléculaire. En effet, la conjugaison bactérienne mise en évidence en 1945 par Joshua Lederberg [25] dépend de l'expression de gènes codés sur ce plasmide.

Le plasmide F code pour sa transmissibilité d'une bactérie dite donneuse à une autre dite receveuse. Ceci fut la preuve d'un brassage génétique et d'une complexité bien plus grande chez la bactérie, ce qui provoqua un regain d'intérêt pour la microbiologie bactérienne.

Au delà de ces fonctionnalités remarquables qui ont permis sa découverte, le F régule sa propre réplication et maintient un très petit nombre de copies dans chaque cellule. De plus, afin de maintenir au mieux le plasmide au sein de la population, il contient aussi un système de partitionnement qui assure à chacune des bactéries filles lors de la division cellulaire de posséder une copie du plasmide.

Afin de simplifier les constructions de biologie moléculaire, nous avons fait le choix de travailler avec le plasmide pZC320 [26] ( $\sim 8\text{kb}$ ) qui est un plasmide F dont la majeure partie des fonctionnalités (la conjugaison ou encore la recombinaison) ont été supprimées. Les fonctionnalités majeures restantes sont le contrôle de sa réplication et le système de partitionnement que nous allons voir plus en détail dans la suite. Dans la suite du manuscrit, ce plasmide sera appelé "miniF".

#### **Les bactéries Hfr**

Le plasmide F est en fait un épisome, c'est-à-dire qu'il peut s'insérer dans le chromosome bactérien. Les protéines nécessaires à ce mécanisme sont codées sur le plasmide même. L'insertion se produit par recombinaison homologue (voir 5.2) dans le chromosome bactérien pour en faire des bactéries à fort taux de recombinaison dite Hfr.

Il est alors possible à la bactérie d'injecter non seulement l'épisome, mais aussi tout le chromosome bactérien à une autre bactérie réceptrice. Ces échanges d'informations entre bactéries constituent ce que l'on appelle le transfert horizontal et permet au niveau d'une population non seulement une propagation plus rapide (il faut l'entendre au sens de l'évolution) d'un gène conférant un avantage sélectif, mais permet aussi à la population d'essayer différentes combinaisons de gènes avec différents avantages.

### **4.1.1 La régulation de la réplication du miniF**

Le plasmide miniF est en fait un dérivé du plasmide F discuté plus haut. Il s'agit donc d'un plasmide ayant sa propre origine de réplication OriS. Ce miniF régule sa réplication

à l'aide de la protéine RepE qu'il encode et de la séquence *incC*.

La fonction de RepE est désormais bien comprise. Elle se lie à la répétition d'une séquence spécifique appelée DR au sein de l'origine de réplication OriS. Cet accrochage provoque l'ouverture de l'ADN double brin et permet la formation d'une fourche de réplication. Cependant, RepE peut aussi former des dimères, qui vont inhiber la transcription de *repE*. Ainsi, il y a une certaine production maximum de RepE dans la bactérie.

En ce qui concerne *incC*, il existe des évidences sur la capacité de liaison de RepE avec *incC* via les mêmes séquences DR, mais le mécanisme et l'effet d'une telle interaction n'ont pas encore été explicités. Il est admis que la séquence *incC* confère l'incompatibilité<sup>1</sup> entre les plasmides de même type, mais les mécanismes restent assez flous. Un mécanisme a été proposé par Uga et al. [27] qui permettrait de rendre compte des différents phénomènes observés. Il est considéré dans ce modèle que RepE en trop grande quantité se lierait à la fois à l'OriS et à *incC* et que les monomères liés de RepE reformeraient des dimères, ce qui aurait pour effet, l'inhibition de la réplication par encombrement stérique. Il ne s'agit ici que d'un modèle auquel aucune preuve directe n'a été apportée.

### 4.1.2 Le système de partitionnement du pZC320

Le pZC320 contient un système de partitionnement constitué par les trois séquences *sopA*, *sopB* et *sopC*. La protéine SopA peut s'autoassembler pour former de longs filaments. Sa polymérisation est initiée par SopB qui est lui même lié à *sopC*. Ainsi, les filaments de SopA entre deux sites *sopC* croissent et séparent les plasmides contenant chacun un de ces sites *sopC* [28].

## 4.2 Plasmide R1

Le plasmide R1 fait partie de la grande classe des plasmides induisant la résistance multiple à un grand nombre de d'antibiotique. Dans les Années 50, des souches multi-résistantes causaient un grand nombre de problèmes aux Japon. Watanabe et Fukasawa [29] mirent en évidence que ces résistances étaient portées par ce type de plasmide.

---

<sup>1</sup>Deux plasmides différents sont incompatibles s'ils ne peuvent coexister au sein du même hôte. Généralement, les plasmides incompatibles bénéficient de système de régulation similaire ou faisant intervenir des espèces biochimiques identiques. On peut alors classer les plasmides en groupes d'incompatibilité.

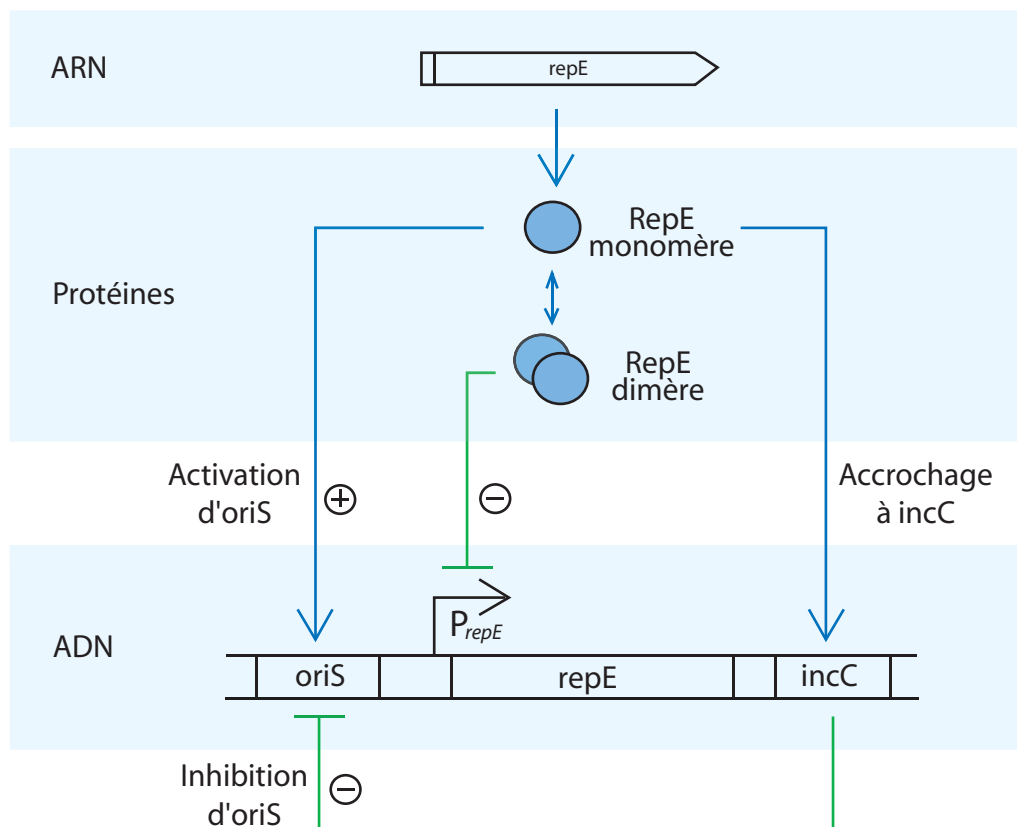


FIG. 4.1: Mécanisme de régulation de la réplication du plasmide F par l'intermédiaire de *repE* et de *incC*.



Nous avons utilisé le plasmide pOU82 qui est un dérivé du plasmide R1 dont la majorité des fonctions comme la résistance multiple à certains antibiotiques ou la conjugaison a été supprimée. Cependant, l'origine de répllication et le système de régulation restent le même. De plus, un autre plasmide que nous avons utilisé, le pOAR64, est issu d'un pOU82 dans lequel un système de partition a été inséré. Dans le reste du manuscrit, nous ferons référence au pOU82 et au pOAR64 respectivement par "miniR1par<sup>-</sup>" et "miniR1par<sup>+</sup>".

### 4.2.1 La régulation de la répllication des plasmides "miniR1par<sup>-</sup>" et "miniR1par<sup>+</sup>"

La répllication des miniR1 est limitée par la protéine RepA. Lorsque cette dernière est en nombre suffisant [30], elle initie la répllication en déstabilisant la forme double brin de l'ADN au niveau de l'origine de répllication *oriR1*. Cette protéine est produite par le biais de deux ARNs messagers transcrits à partir de *PcopB* et *PrepA*. Le premier code, en amont de *repA*, aussi pour une séquence désignée par CopT et pour la séquence de la protéine CopB qui inhibe *PrepA* complètement lorsque le plasmide est en nombre suffisant dans la bactérie. Ainsi, très vite, seul *PcopB* n'est à considérer. De plus, un autre ARN CopA est produit à partir du promoteur *PcopA* sur le brin complémentaire de *repA*. CopA peut former avec CopT un complexe qui empêche la traduction de RepA et qui sera clivé par la RnaseIII laissant comme produit un ARN tronqué ne codant que pour CopB. Une caractéristique intéressante de CopA est qu'il a un temps de vie assez court ( $\sim 2$ min) ce qui rend sa concentration proportionnelle à celle du plasmide. Pour résumer, on peut distinguer deux boucles de rétroaction négative :

- La première implique un arn CopA qui inhibe la production RepA en se liant à l'arn produit à partir de *PcopB*. CopA suit la concentration en plasmide du fait que son promoteur est constitutif et qu'elle ait un petit temps de vie.
- La deuxième met en jeu la protéine CopB qui inhibe la production de RepA à partir du promoteur *PrepA* <sup>2</sup>

Ce mécanisme de régulation du nombre de copies de plasmides est perturbé par un promoteur  $\lambda PR$  qui a été inséré en amont du promoteur *PcopB*. Ce promoteur permet la

---

<sup>2</sup>Il semblerait que cette boucle bloque cette voie de production de manière constitutive sauf dans les situations où CopB est très peu abondante. Typiquement, cette situation apparaît lorsque le plasmide vient d'incorporer une bactérie ou que la concentration de plasmide est très basse.

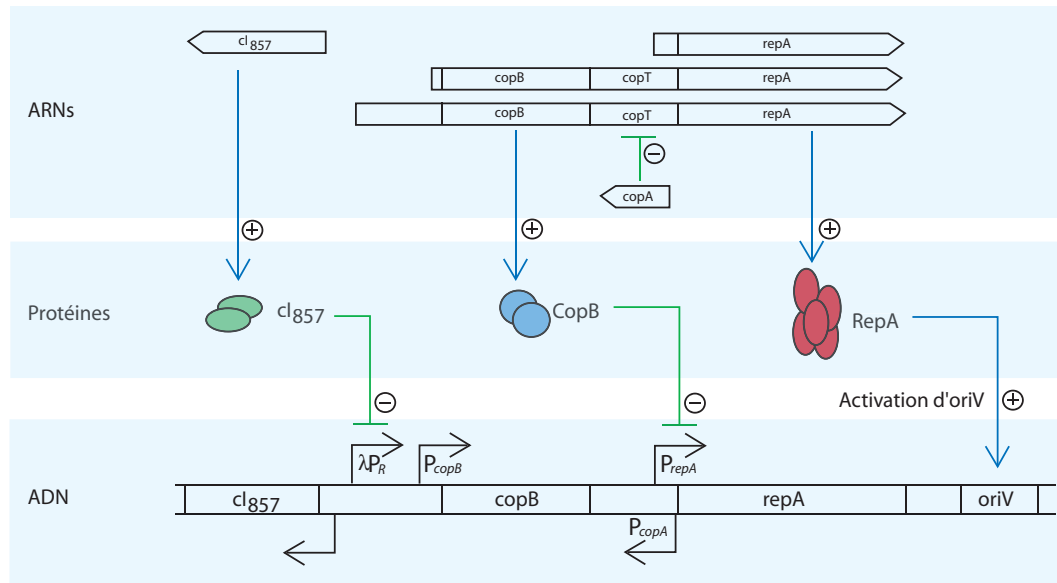


FIG. 4.2: Système de régulation de la réplication du miniR1par<sup>-</sup> et du miniR1par<sup>+</sup>

production d'un arn messager codant aussi pour RepA. Il a aussi été inséré le gène *cI857* exprimé de manière constitutive. Ce gène code pour la protéine CI857 qui se déstabilise à haute température. Ainsi, plus la température est élevée, plus le promoteur  $\lambda P_R$  est exprimé provoquant une augmentation du nombre de copies de plasmides par bactérie avec l'augmentation de la température. Un certain nombre d'expériences [31] [30] ont permis de mettre en avant plusieurs caractéristiques (temps de latence entre deux réplication, action en cis de la protéine RepA) . De plus, quelques modèles [32] [33] [34] de réplication ont déjà été réalisés.

#### 4.2.2 Le système de partition du miniR1par<sup>+</sup>

On retrouve dans le miniR1par<sup>+</sup> les gènes *parA* et *parB* et un site d'accrochage *parS*. Ce système a un fonctionnement similaire au système sopABC. Ainsi, nous ne le détaillons pas.

Toutefois, un phénomène intéressant, dont le mécanisme reste sujet à controverse, apparaît avec ce système de partition. Il a été observé que des gènes aux abords de *parS* pouvaient devenir silencieux et ce quel que soit leur séquence ou leur promoteur [35]. Il ne s'agit pas de répression puisque n'importe quel gène à proximité peut être silencé. Il s'agirait soit de séquestration soit de polymérisation de ParB le long de l'ADN. Cette

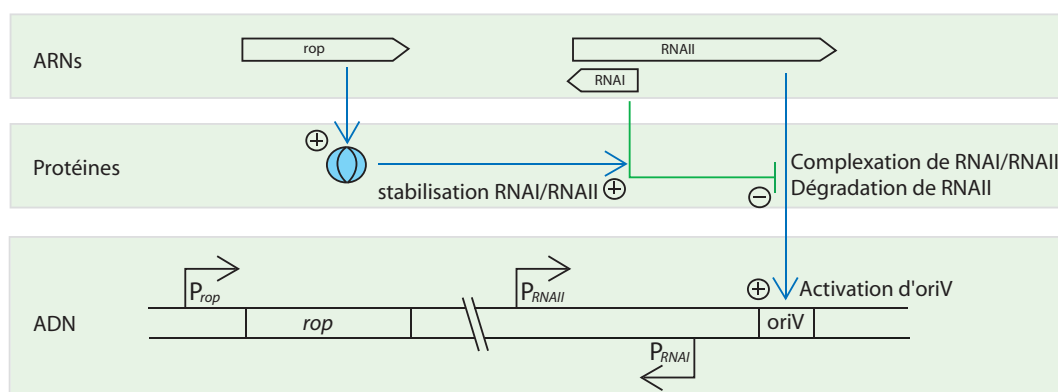


FIG. 4.3: Schéma de régulation de la réplication de colE1

polymérisation empêcherait alors à l'appareillage de transcription d'accéder à l'ADN.

### 4.3 Le Plasmide ColE1

Les plasmides du type ColE1 forment une classe de plasmides très répandus. La plupart des plasmides utilisés pour la production de protéines en font partie (pUC, pET).

Les bactéries portant un plasmide de la classe de ColE1 pouvaient produire des colicines (toxines très efficaces) capables de tuer d'autres organismes. Ce sont des substances qui reconnaissent un récepteur particulier à la membrane et dégradent la membrane. Les bactéries porteuses de ces plasmides pouvaient survivre car elles possédaient un moyen de ne pas être reconnues par ces colicines.

Pour les mêmes raisons de simplicité, nous avons préféré utiliser un plasmide ne codant que pour certaines fonctions essentielles. Nous avons donc choisi le plasmide pBR322. Nous ferons référence à ce plasmide par "mini-colE1" dans le reste du manuscrit.

#### 4.3.1 La régulation du plasmide mini-colE1

La régulation de la réplication du mini-colE1 fait intervenir 2 fragments d'ARN simple brin, RNAI et RNAII. Lorsque RNAII est complètement transcrit, il s'hybride à un site spécifique de l'origine de réplication du mini-colE1, *oriV*, et sert d'amorce à l'ADN polymérase. Ceci permet l'initiation de la réplication. Par contre, RNAI est susceptible de former un complexe avec RNAII. Dans ce cas, la RNase H dégrade le complexe et la RNAII ne peut plus initier la réplication. La complexation des deux ARNs doit avoir lieu

avant que la transcription de RNAII ne soit terminée [36]. On remarque aussi que la protéine Rom (ou Rop selon les sources) qui est encodée sur le plasmide stabilise le complexe RNAI/RNAII [37]. Elle diminue donc le nombre de réplication par cycle cellulaire.

Le mini-colE1 ne contient pas de système de partition. On considère que les différentes copies du plasmide se répartissent de manière binomiale.

Un fait marquant des plasmides tels que le mini-colE1 est que le nombre de copies de plasmides par masse de bactéries augmente avec le temps de division [38]. Si l'hôte est cultivé de manière à ce que son temps de doublement augmente, alors il aura plus de plasmides par bactéries.

Ainsi, plusieurs modèles ont été proposés pour décrire le fonctionnement du système de régulation des plasmides ColE1 dont ceux de Goss et Pecoud [39] [40] ou de Paulsson et Erhenberg [41][42] [43] .

## 4.4 Variabilité du nombre de copies de plasmides

Les plasmides codent pour le réseau biochimique qui régule leur fréquence de réplication. Ce réseau de régulation, différent pour chaque type de plasmide, fixe le nombre de copies de plasmides moyen par bactérie. Un grand nombre d'études se sont attelées à établir le fonctionnement de ces réseaux et à mettre en lumière le rôle de chacun de ses composants. Ces réseaux ne peuvent fixer de manière déterministe le nombre de copies de plasmides au sein de chaque bactérie. En effet, il faut tenir compte des fluctuations du taux de réplication, des variations dans le moment précis de l'initiation de la réplication, ou encore l'allocation de nombre de copies différents aux bactéries filles lors de la division cellulaire. L'ensemble de ces phénomènes font que le nombre de copies de plasmides fluctue entre bactéries au sein même d'une population monoclonale de bactéries.

Plus récemment, les études se sont portées sur les effets du bruit sur les réseaux de régulation du nombre de copies de plasmide et quelques études, en partant de modèles de régulation, établissent la distribution du nombre de copies de plasmides [33] [39] [44]. Cependant, aucune étude expérimentale n'a déterminé la variabilité du nombre de copies de plasmides au sein d'une population de bactéries.

Dans la partie suivante, nous détaillerons le système biologique et le dispositif expérimental que nous avons développés à fin d'étudier le nombre de copies de plasmides au

sein de populations de bactéries.

## Deuxième partie

### Dispositif expérimental

## Chapitre 5

# Les constructions biologiques

Nous voulons obtenir la distribution du nombre de copies au sein de population de bactéries. Expérimentalement, le titrage d'ADN n'offre malheureusement pas une précision suffisante et les PCR quantitative (qPCR) ne donnent accès qu'aux valeurs moyennes lorsqu'elles sont effectuées sur des populations de bactéries. Nous avons donc besoin d'une mesure sur des bactéries individuelles du nombre de copies de plasmides. Une des solutions serait d'utiliser les techniques d'hybridation *in vivo*. Ainsi, Belmont [45] a inséré une succession d'opérateur Lac dans le chromosome d'une bactérie et fait exprimer une fusion lacI-GFP à cette même bactérie. Les fusions LacI-GFP se liant aux opérateurs forment des points focaux de fluorescence au sein de la bactérie. Cependant, cette technique contient deux problèmes majeurs. Premièrement, la diffraction empêche de distinguer un grand nombre de points focaux au sein d'un volume aussi confiné que celui du cytoplasme bactérien. Deuxièmement, Pogliano et al. [46] trouvent 2 à 6 points focaux pour un plasmide ayant un nombre de copies moyen compris entre 5 et 10. Ceci nous indique que soit les plasmides, soit les rapporteurs fluorescents ont tendance à former des agrégats. Dans les deux cas, une mesure du nombre de copies de plasmides par bactérie avec cette méthode mènerait systématiquement à une sous évaluation importante.

Afin d'obtenir le nombre de copies de plasmides, nous avons fait l'hypothèse que le niveau d'expression d'un gène était proportionnel à son nombre de copies. Nous avons alors décidé d'insérer le gène d'un rapporteur fluorescent lié au plasmide. De plus, afin d'obtenir une référence qui serait, elle, indépendante de ce nombre de copies, nous avons en plus cloné un autre rapporteur fluorescent, cette fois-ci dans le chromosome. Cette référence pourrait aussi permettre de s'affranchir du bruit extrinsèque d'expression au

sein de la population [9]. La mesure des deux fluorescences nous donnerait alors accès au nombre de copies de plasmide au sein de chaque bactérie.

## 5.1 L'expression de la protéine fluorescente

Afin d'étudier l'expression génétique au sein de bactéries, une méthode très courante consiste à faire exprimer une protéine fluorescente par la bactérie. Cette production peut se faire de différentes manières. On peut placer l'expression de cette protéine sous le contrôle d'un promoteur similaire à celui d'un gène dont on veut connaître l'expression. Toujours, dans le même registre, on peut faire produire à la bactérie non pas la protéine à étudier, mais la protéine fusionnée<sup>1</sup> avec une protéine fluorescente. Dans notre cas, nous voulions dans un premier temps simplement avoir accès au taux de production des protéines fluorescentes. Nous avons donc opté pour un système où nous avions l'expression de nos protéines fluorescentes sous contrôle d'un promoteur fort et bien caractérisé.

### 5.1.1 Le promoteur PtacI\*

Le promoteur contrôle à la fois la transcription du gène qui lui est attribué et la traduction de l'ARNm produit. Nous avons choisi d'utiliser le PtacI [47] dont la transcription est conditionnée par la présence d'IPTG.

Ce promoteur est très bien réprimé par le gène *lacI<sup>Q1</sup>* [48] exprimé de manière constitutive au sein des souches utilisées. L'ajout d'IPTG dans le milieu de culture lève cette inhibition et permet la transcription du gène en aval du promoteur<sup>2</sup>.

Nous avons fait attention à ne pas modifier la partie régulatrice de l'activité transcriptionnelle.

Toutefois, nous avons plutôt utilisé une autre partie régulatrice concernant la traduction. Ainsi, nous avons utilisé un RBS issu de [49]. Nous appellerons le promoteur ainsi obtenu PtacI\*

---

<sup>1</sup>Typiquement, lors d'une fusion, le gène que l'on veut étudié est modifié de manière à ce que la protéine finale soit liée à la protéine rapporteuse.

<sup>2</sup>En fait, il serait possible d'utiliser aussi du lactose, mais ce dernier est hydrolysé lorsqu'il s'accroche aux LacI alors que l'IPTG ne l'est pas. Utiliser de l'IPTG confère des conditions d'induction constantes au cours du temps.



---

acctgtgaat tctcgagcta gcgttgacaa ttaatcatcg gctcgtataa tgtgtggaat  
 tgtgagcgga taacaatttc acacaggagg taagatctatg

---

La séquence du promoteur ci-dessus montre la région régulatrice de la transcription en italique. La région régulatrice de la traduction est soulignée et le début de la traduction est noté en gras.

**Le gène *lacI*<sup>Q1</sup>** est un "allèle" du gène *lacI*. Ce gène fait partie du système de régulation de l'opéron lactose. L'utilisation de *lacI*<sup>Q1</sup> permet de palier à deux défauts de notre système d'expression. Premièrement, il n'y a qu'un seul opérateur lac dans le promoteur PtacI\*. De ce fait le promoteur PtacI\* montre une fuite d'expression 20 fois supérieure au promoteur lac classique dans les conditions d'expression habituelle d'un gène *lacI*. Deuxièmement, notre gène d'intérêt peut être en copies multiples. Il est donc possible de voir une baisse de la répression à cause du titrage de LacI. En fait, *lacI*<sup>Q1</sup> est identique à *lacI*, seul le promoteur change. Ainsi, la concentration LacI s'en retrouve augmentée d'un facteur 170 pour atteindre 1700 LacI par bactérie. Ce nombre permet une répression efficace de notre gène d'intérêt, même si celui-ci est sur un plasmide à grand nombre de copies.

### 5.1.2 Le choix des protéines fluorescentes

Nous avons choisi d'utiliser la eGFP [50], la mOrange [51] et la YFP-Venus [52].

La eGFP et la mOrange ont été choisies principalement pour leurs caractéristiques (voir annexe A). Leurs spectres d'émission ne se recouvrent que très peu et leurs brillances sont importantes. Le choix de la mOrange par rapport à d'autres protéines rouges s'explique par sa meilleure excitabilité à 488nm.

Nous avons utilisé la YFP-Venus pour son temps de maturation extrêmement rapide et sa forte brillance. Ces caractéristiques font d'elle une très bonne protéine fluorescente pour toutes les applications en temps réel.

### 5.1.3 La terminaison TR2\*

Le but d'un site de terminaison est d'optimiser l'expression d'une protéine. En effet, en limitant l'usage de la polymérase et du ribosome, on augmente leur disponibilité. Ainsi,

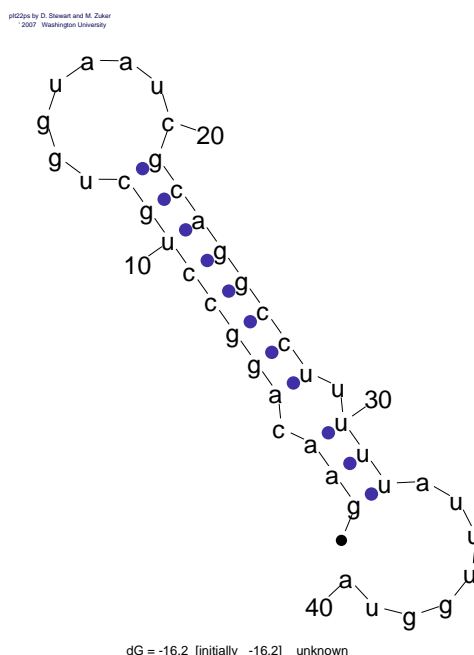


FIG. 5.1: La structure secondaire de la séquence de TR2 a été générée à l'aide du programme m-fold [53] de Zuker. L'arn messenger forme bien une structure en épingle à cheveu qui arrête la transcription. Cette structure est très stable.

nous avons utilisé la séquence de terminaison TR2 étudiée par Gusarov [2] (La première qui vient dans la publication citée).

---

*gaacaggcct* *gctgtaatc* gcaggccttt *ttatttggt* *gacgcatgct* *taagcttact* *ggt*

---

La séquence montrée ne contient pas de codon stop qui sera supposée dans le gène. Les bases en italique s'apparient avec les bases soulignées afin de former une structure en épingle à cheveu (voir figure 5.1). La formation de cette structure décroche l'ARN polymérase, ce qui avorte la transcription au delà du gène d'intérêt. Nous appellerons notre terminaison TR2\*

### 5.1.4 La construction (double PCR)

Lors d'une PCR, on amplifie à partir d'une matrice d'ADN la séquence de celle-ci comprise entre les sites d'hybridation des amorces.

Cependant, il n'y a pas besoin d'une complémentarité stricte entre la séquence nucléo-

tidique et la séquence cible. Seule l'hybridation de l'amorce à l'extrémité 3' est absolument nécessaire. On peut alors insérer des mutations ponctuelles voir même rajouter toute une séquence à l'extrémité 5' de l'amorce <sup>3</sup>. Le fragment amplifié est alors modifié de la même manière.

Nous avons donc utilisé des oligonucléotides s'hybridant sur la séquence à amplifier à leur extrémité 3' et rajoutant les séquences désirées à leur extrémité 5'. Il est déconseillé lors de ce type de PCR d'utiliser des oligonucléotides trop long ( $> 70\text{bp}$ ). Nous avons donc tout d'abord amplifié le gène d'intérêt une première fois en lui ajoutant les 49 dernières bases de *PtacI\** (*PtacI\*\_2*) d'un côté et les 30 premières bases de *TR2\** (*TR2\*\_1*) de l'autre. Nous avons ensuite sur le produit de celui-ci utilisé des oligonucléotides permettant de compléter le *PtacI\** et le *TR2\** (On utilise alors les oligonucléotides *PtacI\*\_1* et *TR2\*\_2*). Ces deux processus correspondent aux étapes 1 et 2 de la figure 5.2

La construction est alors formée du gène d'intérêt (une des protéines fluorescentes) précédé par un promoteur *PtacI\** et suivi d'une terminaison *TR2\**. Les détails se trouvent dans l'annexe B.1.3.

$$\begin{aligned} &PtacI^* - eGFP - TR2^* \\ &PtacI^* - mOrange - TR2^* \\ &PtacI^* - YFP\_Venus - TR2^* \end{aligned} \tag{5.1}$$

## 5.2 Une référence sur le chromosome

Afin d'obtenir un référence d'expression au sein de chaque bactérie, nous avons choisi d'insérer la construction *PtacI\* - eGFP - TR2\** dans le chromosome bactérien. De plus, afin que cette référence ne varie pas en nombre de copies au cours du cycle cellulaire, nous avons choisi de l'insérer à l'opposé de l'*OriC*.

Nous avons utilisé un kit de recombinaison commercial [54] afin d'insérer notre construction à côté du gène *pyrF*<sup>4</sup>. Ce kit nécessite l'utilisation d'une souche naturellement résistante à la streptomycine. Nous avons construit et utilisé la souche Sys qui correspond à

<sup>3</sup>Cette possibilité est à la base des techniques de mutagenèse dirigée par PCR grâce auxquelles il est possible de modifier un gène et la protéine pour laquelle elle code.

<sup>4</sup>Ce gène est placé à l'opposé de l'*oriC*, soit à [1339945-1340682 bp](#)

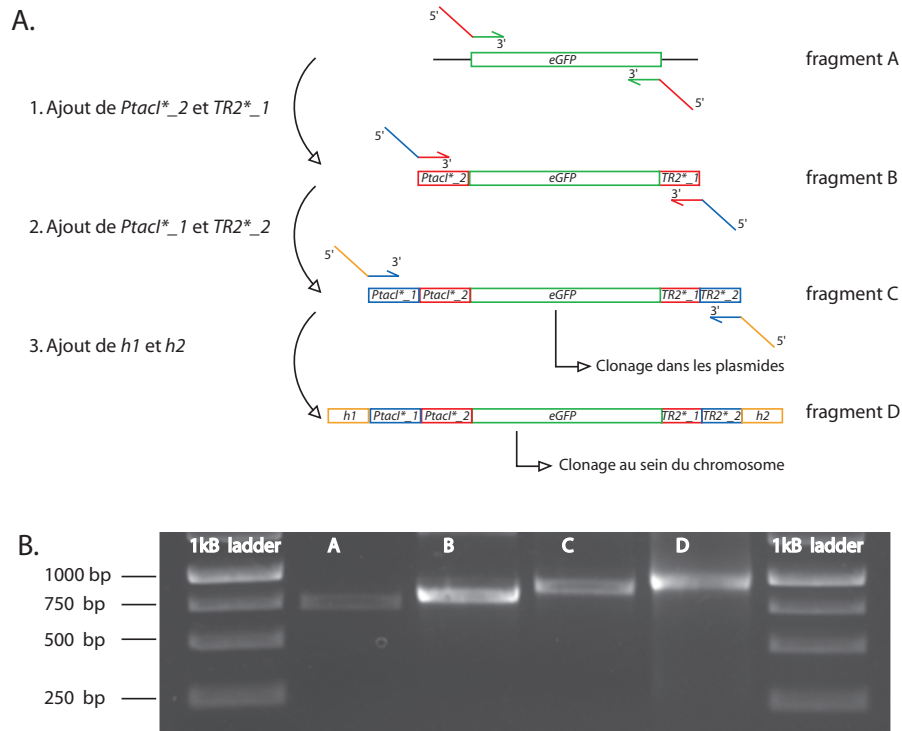


FIG. 5.2: (A) La *eGFP* (fragment A) a été amplifiée plusieurs fois afin d'obtenir les constructions désirées. Pour l'étape 1, 2 et 3, nous avons utilisé respectivement les couples d'amorces *ptacI*-*gfp*/*morV2* et *eGFP*/*xxTomato-TR2\**, *PtacI\** et *TR2\** et *homotac2* et *WG2004-04*. Nous avons alors obtenu respectivement les fragments B, C et D. Il s'agit du même principe pour la *mOrange* et la *YFP\_Venus* (voir annexe B.1.3 pour tous les détails). (B) Gel d'agarose montrant les différents fragments au regard d'un étalon (1kB ladder, Fermentas)

la souche Top10<sup>5</sup> de *Escherichia Coli* dans laquelle le gène *lacI* a été remplacé par son allèle *lacI<sup>q</sup>*.

L'insertion au sein du chromosome se fait par recombinaison homologue [17]. Voici les différentes étapes du kit pour le gène *eGFP* (Voir annexe B.1.3) :

1. Rendre la souche bactérienne compétente pour la recombinaison en le transformant à l'aide du plasmide pSC101-BAD-gbaA (fourni avec le kit). Ce plasmide code pour les protéines nécessaires à la recombinaison comme RecE ou RecT.
2. Amplifier le fragment d'ADN *rpsL-neo* en lui ajoutant les bras d'homologie nécessaire. *rpsL-neo* confère la sensibilité à la streptomycine (*Strep<sup>S</sup>*) et la résistance à la kanamycine (*Kan<sup>R</sup>*). Le produit de cette pcr est *rpsL-neo* flanqué des homologies qui nous intéressent :  $h1 - rpsL - neo - h2$ .
3. Electroporer la souche compétente à l'aide du fragment  $h1 - rpsL - neo - h2$ .
4. Sélectionner les colonies *Strep<sup>S</sup>* et *Kan<sup>R</sup>*. Cette souche sera nommée SysK7.
5. Amplifier la construction en lui ajoutant les bras d'homologie. Pour la construction  $PtacI^* - eGFP - TR2^*$ , cette étape correspond à l'étape 3 du schéma 5.2. On obtient alors la construction  $h1 - PtacI^* - eGFP - TR2^* - h2$ .
6. Electroporer la construction précédente dans la souche SysK7.
7. Sélectionner les colonies *Strep<sup>R</sup>* et *Kan<sup>S</sup>*. Ces bactéries auront remplacé *rpsL-neo* par la construction  $PtacI^* - eGFP - TR2^*$ . On obtient alors la souche SysBRG.

A l'étape 5, nous avons aussi remplacé à l'étape 5 la construction  $PtacI^* - eGFP - TR2^*$  par la construction  $PtacI^* - mOrange - TR2^*$ . Nous avons alors obtenu la souche SysBRO

### 5.3 Le plasmide et sa référence

Les oligonucléotides  $PtacI^*$  et  $TR2^*$  ont été choisi afin de rajouter des sites de restrictions à chacune des extrémités de la construction. Nous avons procédé comme suit (voir annexe B.1.3) :

---

<sup>5</sup>Cette bactérie est issue de la souche K-12 à laquelle quelques modifications ont été incorporés. Une certaine nomenclature permet d'avoir directement accès à ces différentes modifications. La Top10 se présente ainsi : F- *mcrA*  $\Delta(mrr-hsdRMS-mcrBC)$   $\Phi80lacZ\Delta M15$   $\Delta lacX74$  *recA1* *araD139*  $\Delta(ara-leu)$ 7697 *galU* *galK* *rpsL* (*Str<sup>R</sup>*) *endA1* *nupG*

1. Digérer la construction  $PtacI^* - mOrange - TR2^*$  à l'aide des enzymes de restriction nécessaires.
2. Digérer le plasmide avec les mêmes enzymes.
3. Recirculariser le plasmide avec la construction.
4. Transformer chimiquement dans une bactérie DH5 $\alpha$
5. Extraire le plasmide contenant la construction
6. Electroporer ce dernier dans une bactérie contenant la référence sur le chromosome.

Nous avons ainsi obtenu les plasmides pZCBRO, pOUBRO, pOARBRO et pBRBRO respectivement à partir des plasmides pZC320, pOU82, pOAR64 et pBR322. Ces plasmides codent donc pour l'expression de la mOrange. Elles ont été clonées dans des SysBRG qui contiennent la référence d'expression  $PtacI^* - eGFP - TR2^*$  sur le chromosome à côté de *pyrF*. Dans la suite du manuscrit, lorsque nous mentionnerons un plasmide, il s'agira du plasmide dans lequel notre construction a été incorporée.

## Chapitre 6

# Microscopie

La cytométrie de flux permet de mesurer la fluorescence ou d'autres paramètres sur un grand nombre d'objets [55]. Les bactéries défilent dans une zone de détection les unes à la suite des autres. Plusieurs centaines de milliers d'objets peuvent être repérés en quelques minutes. Il existe aujourd'hui des solutions commerciales mais celles-ci restent très onéreuses, et peu versatiles. De plus, elles nécessitent des niveaux d'expression importants.

Une variante de la cytométrie de flux consiste à utiliser un dispositif microfluidique [56] [57] [58]. Ces systèmes consistent à créer des écoulements au sein de canaux de tailles microniques. On effectue ensuite la mesure à un certain endroit du dispositif à l'aide de différents dispositifs optiques. Les écoulements sont créés soit à l'aide de différences de pression[59], soit par électro-osmose[60].

Afin de mesurer la fluorescence de bactéries uniques et d'avoir une statistique suffisante, nous avons développé un système microfluidique largement inspiré de Fu et al. [58].

Les mesures effectuées en cytométrie de flux ne donnent pas accès à l'histoire des bactéries. Nous avons donc mis en place un dispositif de microscopie de fluorescence classique sur lequel nous pouvions suivre la croissance de microcolonies de bactéries. En suivant le lignage des bactéries, nous espérons ainsi suivre l'évolution du nombre copies au cours de la croissance des bactéries.

## 6.1 Le $\mu$ FACS

### 6.1.1 Cellule microfluidique

La cellule microfluidique est constituée de 3 canaux se rejoignant en un même point. Afin de fabriquer les microcanaux dans lesquelles circuleront les bactéries, nous avons utilisé des méthodes usuelles de lithographie douce [61]. Pour cela, nous avons tout d'abord fabriquer des moules dans lesquels sera versé un élastomère (RTV615, Bayer).

Ces moules présentent des reliefs correspondant aux canaux que nous désirons. L'élastomère moulé contient l'empreinte des canaux en négatif. On peut alors coller l'élastomère à une lamelle en verre ou en quartz, ce qui achève la fabrication de la cellule (Voir annexe B.2). Dans le cas d'une lamelle en quartz, nous faisons en sorte à ce que le scellement ne soit pas trop robuste afin de récupérer la lamelle à la fin de l'expérience.

Une fois rempli d'une solution ionique contenant les bactéries, il est possible de créer un écoulement soit en appliquant un champ électrique ( $\sim 40 \text{ V}\cdot\text{cm}^{-1}$ ) soit une différence de pression. Notre système de détection a une petite profondeur de champ et nous devons utiliser des canaux qui ne font que  $2\mu\text{m}$  de hauteur. En appliquant un champ, nous avons remarqué que les bactéries adhéraient systématiquement à la paroi et bouchaient le canal en à peine quelques secondes. Pour empêcher cela, il faudrait traiter les parois du canal. Cela peut se faire à l'aide d'une solution concentrée de PDMA. Ces solutions sont suffisamment visqueuses pour que nous ayons besoin d'appliquer une pression supplémentaire pour réussir à remplir nos canaux. On doit alors fortement sceller les lamelles et l'élastomère. Par opposition, simplement remplir un des puits sans même contrôler les différences de pression nous permettait d'avoir des écoulements à peu près stables pendant plusieurs dizaines de minutes avec des débits qui nous permettaient de mesurer au moins une dizaine de bactéries par secondes. Dans cette configuration nous ne savons pas si l'écoulement est créé par une différence de pression et/ou par capillarité et évaporation. Nous avons cependant opté pour cette dernière méthode dans un premier temps.

### 6.1.2 Le microscope inversé

La cellule microfluidique est observée à l'aide d'un microscope inversé (Olympus IX81). La cellule est placée sur une platine motorisée (Marzhauser) et nous la visualisons avec



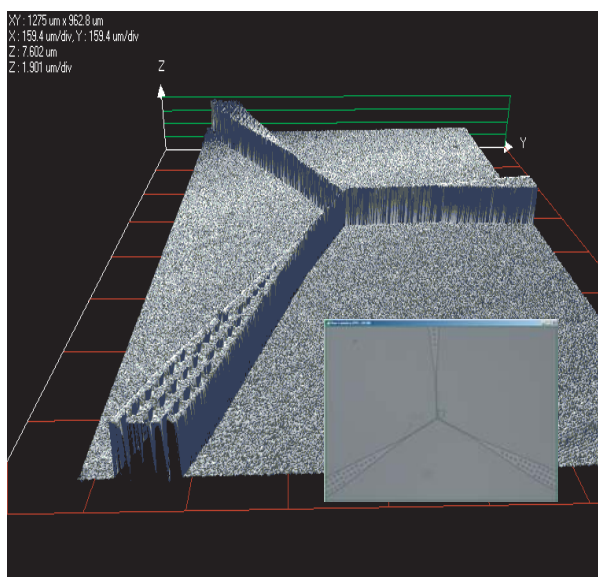


FIG. 6.1: Les canaux formant un Y ont été contrôlés à l'aide d'un profilomètre optique. Si sur cette image, les canaux font  $5\mu\text{m}$  de hauteur, nous avons par la suite modifié le protocole afin qu'ils n'en fassent plus que 2.

un objectif à immersion à huile  $63\times$  d'ouverture numérique 1,4 (Zeiss). L'éclairage peut soit simplement être l'éclairage en lumière blanche du microscope, soit un faisceau laser modulé spatialement en phase. On reforme ensuite l'image des canaux soit à l'infini avec les oculaires, ou sur une caméra CCD refroidie (Pentamax, Princeton Instruments.) ou sur un duo de photodiodes avalanches (APDs, EG&G). Dans ce dernier cas, compte tenu de l'optique et de la taille des détecteurs, nous avons un champ d'observation circulaire d'environ  $10\mu\text{m}$  de diamètre.

Le laser à 488nm (Coherent, Sapphire 488-200) permet une excitation de fluorescence importante. Nous modifions son faisceau à l'aide d'un modulateur spatial de phase ([LCR-720](#), [Holoeye](#)). Ce dernier permet d'éclairer de manière homogène une zone restreinte<sup>1</sup>.

<sup>1</sup>Typiquement, il suffit de fournir à un logiciel le motif désiré (binaire seulement pour notre module). Ce logiciel effectue la transformée de fourrier inverse de l'image (en phase) et la transfère sur le LCR-720. Ainsi, lorsque l'on se place dans le plan de fourrier (pour ce qui nous intéresse, dans le plan focal de l'objectif), la diffraction fait que nous reformons de nouveau le motif fourni au logiciel. En pratique, il faut tenir compte de deux phénomènes : Le premier est qu'il existe toujours un ordre 0 non modifié, ce qui crée un point focal intense dans le plan de fourrier. Le deuxième est qu'il existe toujours un "fantôme", réplique symétrique par rapport à l'ordre 0. Il est possible d'optimiser la diffraction vers l'ordre qui nous intéresse en jouant sur la polarisation du faisceau. Nous avons donc placé une lame  $\lambda/2$  avant le module.

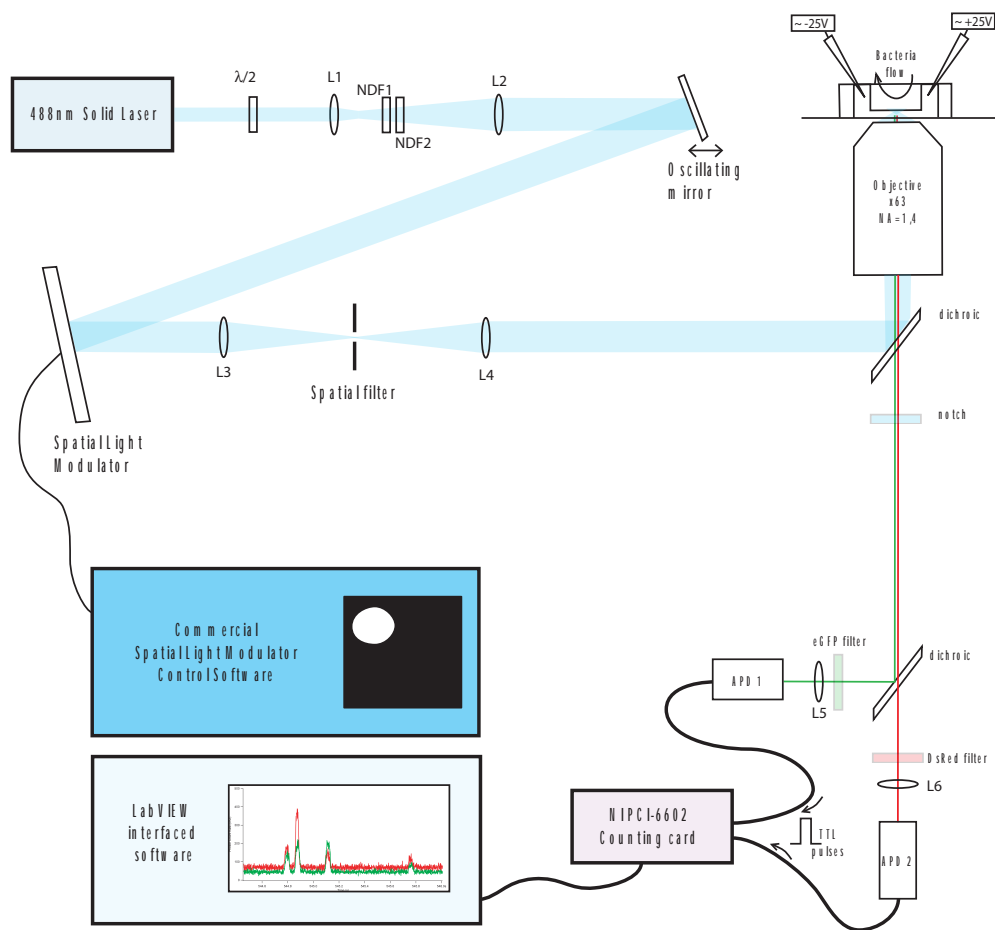


FIG. 6.2: Au sein de ce montage, seul la voie des APDs est montrée par souci de clarté.

Par contre, nous avons pu observer un important speckle directement lié au module optoélectronique. Nous avons choisi de déplacer le faisceau laser en plaçant un miroir sur un module oscillant, ce qui n'influence pas à priori la zone d'excitation, mais qui permet de perdre suffisamment de cohérence spatiale pour éliminer le speckle.

Lorsque nous désirons effectuer une mesure de la fluorescence sur un grand nombre de bactéries, nous utilisons l'excitation laser et mesurons la fluorescence à l'aide des APDs. Les APDs émettent un pulse TTL pour chaque photon reçu, et nous comptons ces pulses à l'aide d'une carte de comptage (NI PCI-6602, National Instruments). Lorsqu'une bactérie passe dans le champ d'observation des APDs, nous mesurons un incrément de la fluorescence qui dure le temps que la bactérie passe dans le champ d'observation.

L'ensemble des instruments est contrôlé par une interface LabView que nous avons développée. Les traces sont alors analysées ultérieurement (voir annexe C) et la fluorescence totale de chaque bactérie peut-être extraite.

## 6.2 Le Timelapse

Afin d'avoir accès au lignage des bactéries, nous les observons sous un microscope droit (Olympus BX51).

Les bactéries sont tout d'abord étalées sur du LB-agar avec les antibiotiques et l'inducteur si nécessaire. On accole ensuite une lamelle de verre afin de les contraindre dans un plan. Nous les imageons ensuite en contraste de phase à l'aide d'un objectif 100× d'ouverture numérique 1,3 (Olympus). La fluorescence est excitée à l'aide d'une lampe à vapeur de mercure et nous reformons l'image sur une CCD refroidie (Micromax, Princeton Instruments). L'ensemble est thermalisé à 37°C. La lame est montée sur une platine motorisée et nous contrôlons la prise d'images avec une interface Labview que nous avons développée. Nous suivons ensuite plusieurs colonies au cours du temps (typiquement 5 ou 6 en prenant une image toutes les 3'). Les images sont ensuite analysées sous Matlab à l'aide du [script](#) de Michael Elowitz.

---

De plus, il est possible de se débarrasser complètement de ces deux phénomènes en appliquant un filtre spatial dans le plan de fourrier.

## Troisième partie

### Résultats

# Chapitre 7

## Une première caractérisation

### 7.1 Temps de division et plasmides

Il est nécessaire de savoir à quel point les bactéries sont affectées par la présence des plasmides au sein de leur organisme. De plus, nous voulions savoir si la production des protéines fluorescentes induisait un poids métabolique trop grand ou non. En mesurant l'évolution de la densité optique à 600nm ( $OD_{600}$ ) d'une culture liquide de bactéries en fonction du temps, il est possible d'extraire le temps de doublement<sup>1</sup> (Voir figure 7.1).

Nous avons obtenu les temps de doublement des souches que nous avons construites à 37°C. Les souches contenant les plasmides n'ont pas de temps de division significativement

Souche	Temps de Division
SysBRG	$29.6 \pm 0.6$
SysBRO	$30.6 \pm 1.0$
SysBRG + miniF	$28.4 \pm 0.7$
SysBRG + miniR1par <sup>-</sup>	$30.0 \pm 0.3$
SysBRG + miniR1par <sup>+</sup>	$29.9 \pm 0.9$
SysBRG + miniColE1	$29.8 \pm 1.2$

différent de celle ne contenant aucun plasmide. Toutefois, en conclure que les plasmides

---

<sup>1</sup>Les bactéries au sein d'une culture croissent de manière exponentielles mais finissent par saturer la culture. Nous nous plaçons systématiquement dans la phase exponentielle de la croissance. Ainsi, toutes les  $OD_{600}$  étaient inférieures à 0,2 afin de s'assurer qu'aucun ralentissement de la croissance n'affecte notre mesure.

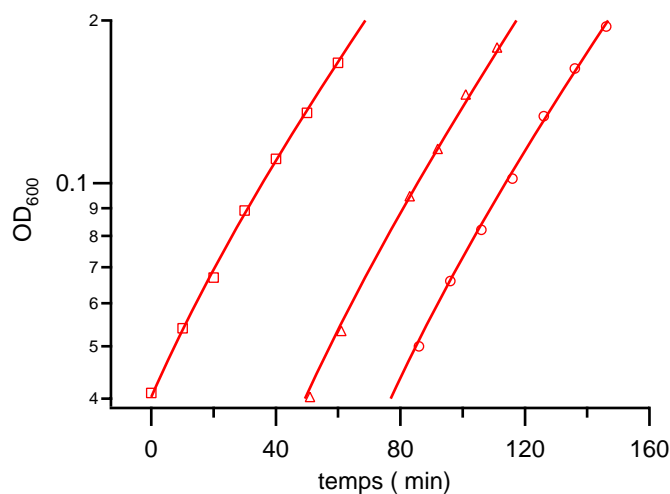


FIG. 7.1:  $OD_{600}$  en fonction du temps. Sont représentées par des symboles différents trois croissances avec la même souche à  $37^{\circ}\text{C}$  dont l'origine des temps est arbitraire. En trait plein sont les ajustements par une exponentielle de chacun d'eux.

n'induisent pas de poids métabolique n'est pas possible. La mesure du temps de division ne reste qu'un indicateur de la "forme" des bactéries.

Quoiqu'il en soit, on peut noter que le temps de division à  $37^{\circ}\text{C}$  est inférieur au temps nécessaire à la réplication du chromosome qui est d'environ 40 minutes. Dans ce cas, il y a 4,70 origines de réplication du chromosome, 1,74 terminii de chromosome et 5,92 fourches de réplication en moyenne par bactérie[62].

## 7.2 Temps d'induction

Nous avons voulu savoir comment évoluait la production de la protéine fluorescente avec le temps pendant lequel nous induisons. Pour cela, nous avons mesuré sur plusieurs populations leur fluorescence. Plutôt que de répéter nos expériences un certain nombre de fois avec différents temps d'induction, nous avons préparé les souches sysK7+miniColE1-YFP dans les mêmes conditions que pour notre méthode d'induction. Ensuite, nous avons observé la croissance de ces bactéries sur un bloc de LB-agar contenant de l'IPTG à  $1\text{mM}$  (voir figure 7.2).

La fluorescence moyenne des bactéries au sein de la colonie augmente pour saturer avec un temps caractéristique de l'ordre de l'heure.

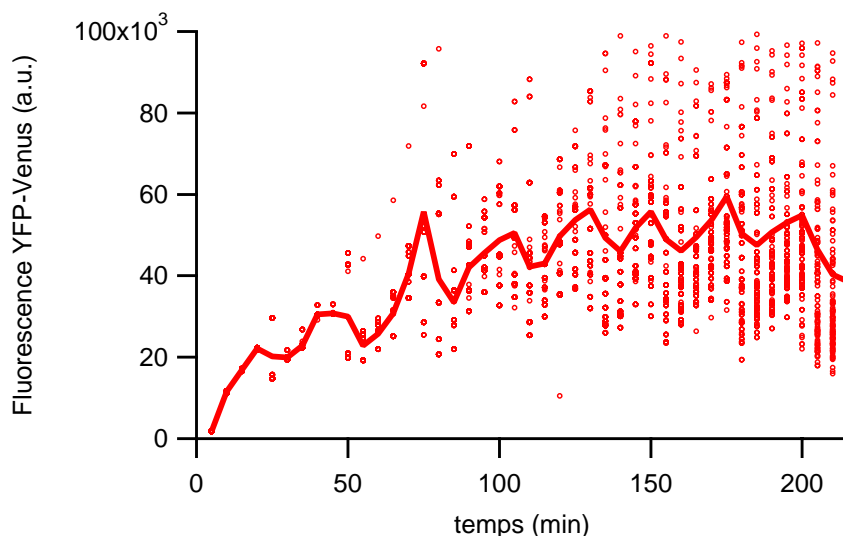


FIG. 7.2: Evolution de la fluorescence avec le temps au sein d'une colonie de bactéries. Les cercles sont les fluorescences totales de chacune des bactéries. La ligne rouge est la fluorescence moyenne des bactéries.

Dans ces conditions, les temps de division des bactéries que nous avons obtenus ne sont pas constants au cours du temps. Plus précisément, les temps de division sont un peu plus longs au début. Il est probable que le transfert des bactéries sur le bloc de LB-agar et le temps nécessaire à la thermalisation du système provoquent des temps de division plus grand au début de l'enregistrement. De plus, la fluorescence au bout d'un temps supérieur à 200 minutes décroît légèrement ce qui reflète un effet de photoblanchiment du à une exposition trop longue à la lumière d'excitation et/ou une diminution d'IPTG dans le milieu de culture<sup>2</sup>.

Malgré ces quelques bémols, il semblerait que prendre un temps d'induction de l'ordre de 60 minutes soit un compromis correct entre le niveau de fluorescence maximum et un temps d'induction le plus court possible.

---

<sup>2</sup>La diminution d'IPTG ne serait liée qu'à son internalisation par une quantité croissante de bactéries car l'IPTG n'est pas dégradé. Cependant, au vu du volume d'une bactérie ( $\sim 10^{-15}$ L) et celui de l'échantillon ( $\sim 25\mu\text{L} = 25 \cdot 10^{-6}\text{L}$ ), il semble improbable que ce soit le cas.

### 7.3 Concentration d'IPTG

Nous avons mesuré la fluorescence moyenne de la même population en l'induisant à différentes concentrations d'IPTG (voir figure 7.3). Entre 0,1 et 10 mM d'IPTG, La fluorescence moyenne de la population ne varie que très peu. Glascock et Weickert [48] donnent une concentration de saturation entre 10 et 30  $\mu$ M pour un système équivalent au notre.

Nous avons observé une fluorescence moyenne légèrement plus faible à 0,1 mM et préférons nous placer un ordre de grandeur au dessus afin d'être sûr de nous affranchir des erreurs liées aux volumes pipetés. De plus, nous voulions pouvoir observer d'autres promoteurs comme *Lac* sans pour autant changer les concentrations d'induction. Toujours selon Glascock et Weickert [48], le promoteur *Lac* lui présente une saturation plutôt à 100 mM. Nous nous placerons par la suite à 1 mM d'IPTG pour l'induction.

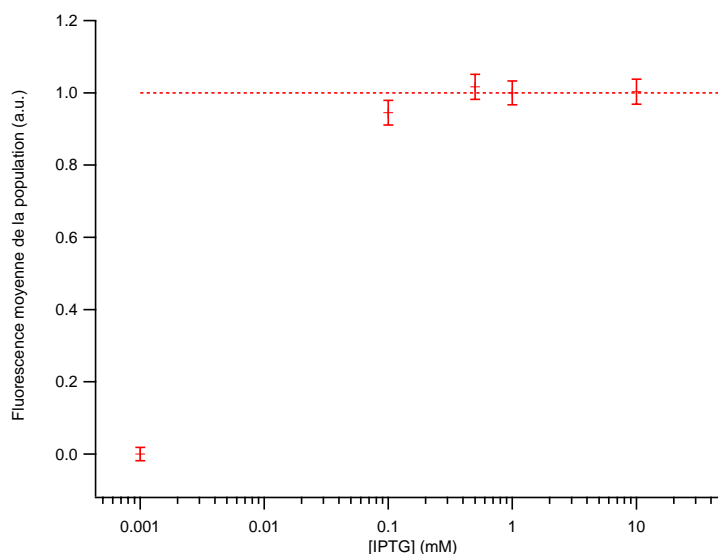


FIG. 7.3: La fluorescence moyenne ne varie pas de manière significative entre 0,1 et 10 mM. La droite en pointillés marque la valeur à 1 mM d'IPTG.

### 7.4 Variabilité inter colonies et OD d'induction

Nous avons testé plusieurs colonies induites à des  $OD_{600}$  comprises entre 0,18 et 0,25. Les valeurs obtenues pour les moyennes et les variances ne sont pas discernables. Les



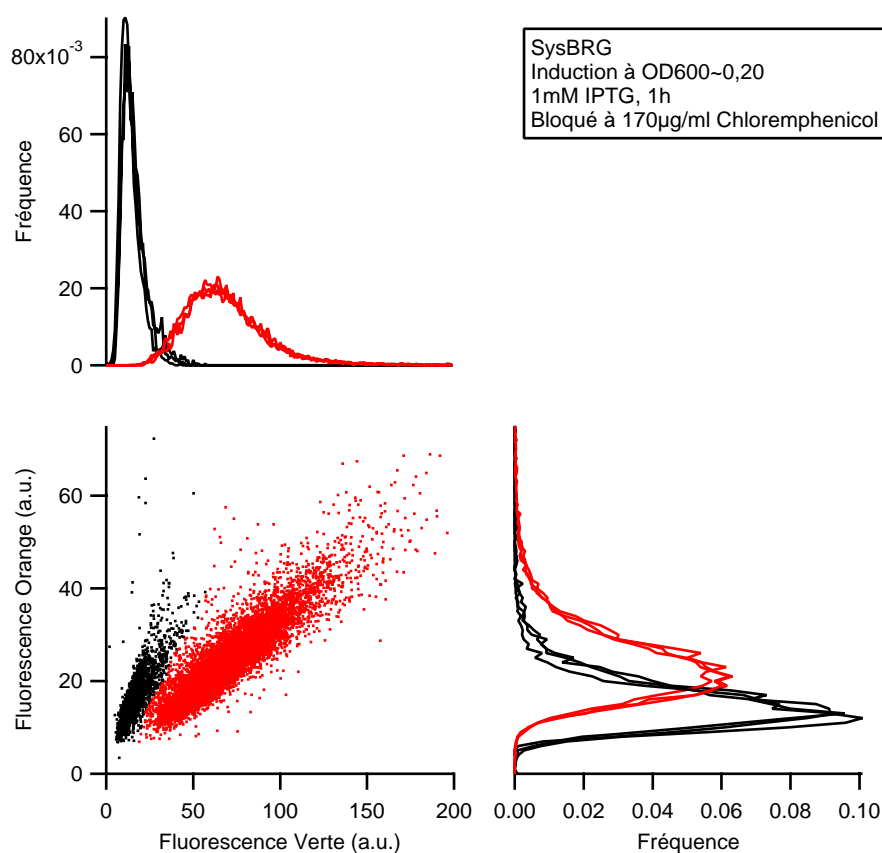


FIG. 7.4: Sont représentés le diagramme de dispersion et les histogrammes de fluorescences correspondants à trois colonies de SysBRG issus d'une même boîte de pétri. En noir les cultures non induites, ce qui traduit l'autofluorescence des bactéries. et en rouge les cultures induites à 1mM IPTG.

histogrammes sont pratiquement indiscernables pour 3 colonies différentes (voir figure 7.4).

## Chapitre 8

# Estimation du nombre de copies moyen pour chacun des plasmides

### 8.1 PCR en temps réel

Nous avons quantifié le nombre de copies (PCN) moyen de plasmides par bactéries par PCR en temps réel. Cette méthode a pour avantage d'être rapide et de s'affranchir du nombre de bactéries utilisées pour la quantification [63]. En effet, nous avons utilisé trois couples d'oligonucléotides qui permettent d'amplifier des fragments d'une centaine de bases au sein des gènes *DnaA*, *pyrF* et *Bla*.

On mesure l'évolution des différentes quantités d'ADN au cours de l'amplification à l'aide d'un intercalant SYBR-green. Les mesures ont été effectuées avec un [LightCycler 1.5](#) (Roche Applied Science). Leurs évolutions permettent de connaître les concentrations initiales relatives entre les différents fragments ciblés.

Le gène *DnaA* est situé au niveau de l'origine de réplication du chromosome et le gène *pyrF* à l'opposée de *DnaA* sur le chromosome. Le rapport de leurs concentrations devrait donner le nombre moyen d'origines de réplication du chromosome par terminus au sein de la population. Bremer et al. [62] ont mesuré, pour un temps de division de 30 minutes, environ 2,7 origines par terminus.

Le gène *Bla* est celui qui code pour la résistance à l'ampicilline. Il est porté par tous les plasmides que nous avons étudiés. Lorsqu'on comparera la concentration de *Bla* à celui de *pyrF*, nous espérons obtenir le nombre de plasmides par terminus. Nous n'avons pas utilisé les gènes *eGFP* ni *mOrange* car nous voulions pouvoir faire les PCR sur toutes les souches que nous possédions.

Les PCRs quantitatives que nous avons faites sur les lysats de bactéries ou sur les bactéries entières n'ont jamais donné de résultats cohérents. Nous avons donc d'abord extrait l'ADN des bactéries à l'aide d'un kit d'extraction d'ADN génomique [QIAamp](#). C'est sur l'ensemble de l'ADN élué que nous avons fait les PCR en temps réel.

	mini-F	mini-R1-par <sup>-</sup>	mini-R1-par <sup>+</sup>	mini-ColE1
$[DnaA]/[pyrF]$	2, 9	2, 3	2, 6	2, 4
$[Bla]/[pyrF]$	0, 5	3, 2	3, 8	23, 4

Premièrement, le nombre d'origines de réplication du chromosome par terminii est cohérent avec Bremmer et Al. [62].

Pour ce qui est du mini-F, le rapport  $[Bla]/[pyrF]$  donnerait un nombre de mini-F moyen inférieur à 1. Cela nous indique que l'extraction d'ADN que nous avons faite avant les PCR quantitatives a probablement un rendement supérieur pour les fragments d'ADN plus grands<sup>1</sup>.

Par contre, si on suppose un rendement identique pour tous les plasmides, on devrait pouvoir comparer les plasmides entre eux. Par exemple, si on fixe le nombre de copies moyen du mini-F à 1, 4 [26], on obtient respectivement 8, 9, 10, 6 et 65, 5 pour les nombres de copies moyen du miniR1par<sup>-</sup>, miniR1par<sup>+</sup> et du miniColE1 ce qui donne les bons ordres de grandeur par rapport à la littérature [44].

---

<sup>1</sup>Le kit de purification QIAamp est généralement utilisé pour purifier des fragments d'ADN allant jusqu'à 50kb. L'ADN chromosomique qui contient *DnaA* et *pyrF* était toujours fragmenté en morceaux plus petits. Les plasmides que nous avons utilisés font moins de 8kB.

## 8.2 Calibration des protéines fluorescentes

Une fois le système de mesure mis en place, il est nécessaire de calibrer la détection de la mOrange par rapport à celle de l'eGFP. En effet, non seulement les caractéristiques des rapporteurs fluorescents sont différentes (excitation et émission en longueur d'onde et en intensité), mais le chemin optique nécessaire à la détection est aussi différent pour chacune des protéines (les caractéristiques spectrales du dichroïque et des filtres passe-bande sur les deux voies sont différentes.). Bien qu'il soit en principe possible de calculer les intensités relatives détectées pour nos deux rapporteurs<sup>2</sup>, certaines des grandeurs nécessaires ne sont pas assez précises. Nous avons alors opté pour une calibration expérimentale.

### 8.2.1 Correction de la fluorescence pour les fluorescences moyennes.

Le signal de fluorescence de chaque bactérie a plusieurs composantes. Ainsi,

$$\begin{aligned} F_V &= P_V + A_V + \alpha P_O \\ F_O &= a P_O + A_O + \beta P_V \end{aligned}$$

$F_V$ ,  $P_V$  et  $A_V$  sont respectivement la fluorescence détectée sur le canal vert, la fluorescence produite par les protéines eGFP et l'autofluorescence verte pour une des bactéries de la population. De même  $F_O$ ,  $P_O$  et  $A_O$  sont respectivement la fluorescence détectée sur le canal orange, la fluorescence produite par les protéines mOrange et l'autofluorescence orange pour une des bactéries de la population.  $\alpha$  et  $\beta$  sont respectivement les coefficients de fuite de la fluorescence de la mOrange vers le canal vert et de la fluorescence de la eGFP vers le canal orange.  $a$  est un coefficient que nous injectons "artificiellement". Il nous permettra de normaliser la fluorescence produite par la mOrange par rapport à la fluorescence produite par la eGFP.

Nous voudrions obtenir  $P_V$  et  $P_O$ . Il n'est pas possible d'obtenir au sein de chaque bactérie la composante d'autofluorescence. Cependant, si on effectue deux mesures, l'une dans laquelle les bactéries ont été induites et l'autre dans laquelle elles ne l'ont pas été, il

---

<sup>2</sup>Nous connaissons toutes les caractéristiques de l'eGFP et de la mOrange et la transmission des filtres et miroirs dichroïques sont fournies par le constructeur.

est possible d'obtenir  $\langle F_V \rangle$ ,  $\langle A_V \rangle$ ,  $\langle F_O \rangle$  et  $\langle A_O \rangle$ . On trouve alors :

$$\langle P_V \rangle = \frac{a}{a - \alpha\beta} \{ (\langle F_V \rangle - \langle A_V \rangle) - \frac{\alpha}{a} (\langle F_O \rangle - \langle A_O \rangle) \} \quad (8.1)$$

$$\langle P_O \rangle = \frac{1}{a - \alpha\beta} \{ (\langle F_O \rangle - \langle A_O \rangle) - \beta (\langle F_V \rangle - \langle A_V \rangle) \} \quad (8.2)$$

On pourra donc obtenir  $\langle P_V \rangle$  et  $\langle P_O \rangle$  dans un premier temps.

### 8.2.2 Equivalence entre eGFP et mOrange et estimation de la fuite

Nous avons comparé le niveau moyen de fluorescence verte au sein de la souche SysBRG et le niveau moyen de fluorescence orange au sein de la souche SysBRO. La souche SysBRG est notre souche de départ (Top10) dans lequel nous avons intégré sur le chromosome la construction permettant d'exprimer l'eGFP. Ainsi, elle ne contient pas le gène de la mOrange et il ne peut donc pas y avoir de fuite de la protéine mOrange vers le canal de détection vert pour la souche SysBRG. Nous avons construit la souche SysBRO de la même manière que la souche SysBRG à l'exception du fait que nous avons remplacé le gène de l'eGFP par la mOrange. La souche SysBRO ne contient donc pas de protéine eGFP et donc pas de fuite de fluorescence de la mOrange vers le canal de détection vert. Lorsqu'on réécrit les équations 8.1 pour les sysBRG et 8.2 pour les sysBRO, nous obtenons donc :

$$\begin{aligned} P_V(SysBRG) &= F_V(SysBRG) - A_V(SysBRG) \\ P_O(sysBRO) &= \frac{1}{a} (F_O(SysBRO) - A_O(SysBRO)) \end{aligned}$$

Fixer  $\langle P_V(SysBRG) \rangle = \langle P_O(sysBRO) \rangle$  nous permet d'obtenir le coefficient  $a = \frac{14,2 \pm 0,9}{24,2 \pm 0,8} = 0,58 \pm 0,04$ . Ceci nous permettra de comparer l'expression liée au chromosome et à celle liée au plasmide au sein d'une même souche.

De plus, dans les souches ne contenant que la eGFP, l'incrément de fluorescence sur le canal de détection de la mOrange n'est dû qu'à la composante "orange" du spectre d'émission de fluorescence de la eGFP (figure 8.1).

Nous avons ainsi estimé la fuite de fluorescence de la eGFP dans le canal orange à  $\beta = 0,1674 \pm 0,0005$ . Le fort coefficient  $\beta$  reflète le compromis fait lors du choix du couple de rapporteurs fluorescents (voir annexe A).

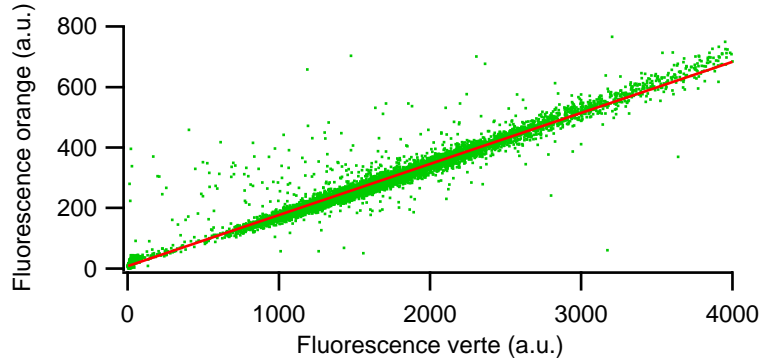


FIG. 8.1: Diagramme de dispersion de la souche SysK7 contenant le plasmide pBRBRG. Les axes ne sont pas à la même échelle. Nous obtenons un coefficient de corrélation  $(\langle \Delta F_V \Delta F_O \rangle / \sqrt{\langle \Delta F_V^2 \rangle \langle \Delta F_O^2 \rangle})$  entre la fluorescence verte et la fluorescence orange de 0,972. L'analyse des intensité moyennes de la population donne un coefficient de fuite  $\beta = 0,1674$ . La droite rouge correspond à un ajustement linéaire de la fluorescence rouge par rapport à la fluorescence verte de chaque bactérie. Un résultat similaire est obtenu pour la souche sysBRG.

Par contre, nous n'avons pu détecter de fuite de fluorescence de la mOrange dans le canal vert. Ceci est cohérent avec le fait que le spectre de la mOrange ne recoupe pas avec le canal de détection vert (voir annexe A).

### 8.3 Linéarité de l'expression avec le nombre de copies

Afin de vérifier que la fluorescence augmente bien linéairement avec le nombre de copies, nous voulions insérer des multimères de notre construction au sein d'un plasmide. Malheureusement, nous n'avons obtenu qu'une seule construction, celle du plasmide miniF avec un dimère de notre construction avec les promoteurs tête-bêche. Nous avons conclu des expériences qu'aux bas nombres de copies, l'expression est bien linéaire avec le nombre de copies du gène (voir figure 8.2).

### 8.4 Expressions moyennes et PCN moyen

Considérons une bactérie contenant  $n_C$  copies du chromosome et  $n_P$  copies d'un plasmide en moyenne pendant l'induction. Le chromosome contient le gène du rapporteur

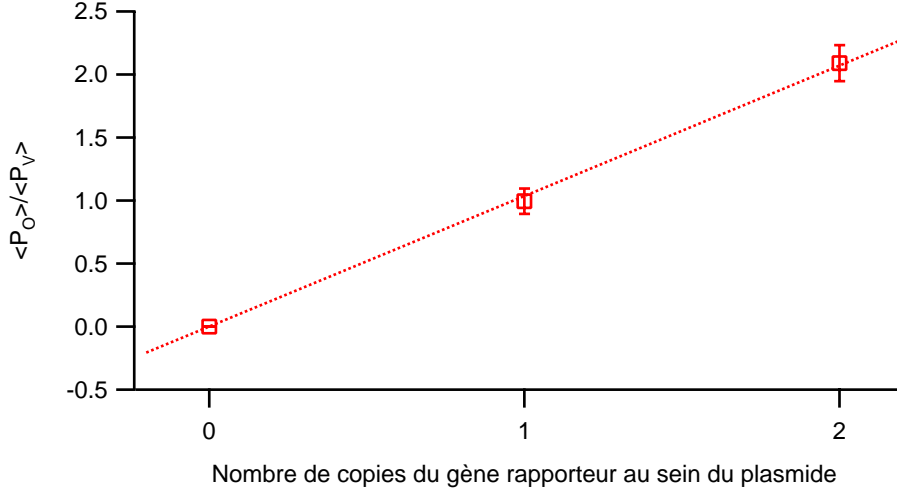


FIG. 8.2: Expression lié au plasmide normalisé par rapport à l'expression liée au chromosome,  $P_O/P_V$ , en fonction du nombre de copies de la construction inséré dans le plasmide. Le zéro ici a été imposé. Les ( $\square$ ) sont les mesures que nous avons obtenus et les barres d'erreurs sont les dispersions que nous avons obtenus pour différentes mesures. En pointillés, un ajustement linéaire toujours en imposant l'ordonnée à l'origine nulle. Nous avons alors une pente  $p = 1,03 \pm 0,02$

fluorescent vert et le plasmide le gène du rapporteur fluorescent orange. La production de protéines eGFP dans une bactérie est la somme des productions de cette protéine par chacune des copies de la construction eGFP au sein de la bactérie. Dans un premier temps, nous proposons d'écrire les expressions des protéines fluorescentes  $P_V$  à partir du chromosome et  $P_O$  à partir des plasmides comme :

$$P_V = \sum_{i=1}^{n_C} P_V^i \quad (8.3)$$

$$P_O = \sum_{i=1}^{n_P} P_O^i \quad (8.4)$$

où les  $P_V^i$  et les  $P_O^i$  sont respectivement les expressions liées à i-ème copie du chromosome et du plasmide. On peut alors poser certaines hypothèses :

- En moyenne, l'expression pour chacune des copies du chromosome est la même. Il paraît difficile d'imaginer qu'il en soit autrement si nous excluons l'éventualité d'une mutation qui affecterait seulement une des copies du gène rapporteur. On peut réécrire cette hypothèse sous la forme :

$\langle P_V^i \rangle = \langle P_V^1 \rangle \forall i$ .  $P_V^1$  est l'expression du gène sur le chromosome pour une seule copie.

- En moyenne, l'expression pour chacune des copies du plasmide est la même. Nous pouvons faire ici la même remarque qu'à l'hypothèse précédente. Donc :

$\langle P_O^i \rangle = \langle P_O^1 \rangle \forall i$ .  $P_O^1$  est l'expression du gène sur le plasmide pour une seule copie.

- En moyenne, l'expression pour une copie du plasmide est la même que pour une copie du chromosome. Cette hypothèse est plus forte. Elle est clairement vérifiée si l'expression d'un gène ne dépend que de son promoteur et de la régulation de ce promoteur. On obtient alors :

$$\langle P_O^1 \rangle = \langle P_V^1 \rangle.$$

- Au sein d'une bactérie, l'expression d'une copie du chromosome  $P_V^i$  ne dépend pas du nombre de chromosome  $n_C$ . Si cela n'était pas, nous n'aurions pas pu avoir de linéarité entre l'expression et le nombre de copies du gène rapporteur dans la partie précédente.
- Au sein d'une bactérie, l'expression d'une copie du plasmide  $P_O^i$  ne dépend pas du nombre de plasmide  $n_P$ . On peut faire la même remarque que dans l'hypothèse précédente. Par contre, il faut noter que la gamme de nombre de copies peut être fortement différentes.

Ici, "en moyenne" désigne une moyenne sur toutes les bactéries d'une population. Il est alors facile d'écrire  $\langle P_V \rangle = \sum_{n_C} p(n_C) \sum_{i=1}^{n_C} \langle P_V^i \rangle = \langle P_V^1 \rangle \sum_{n_C} p(n_C) n_C$  où  $n_C$  est le nombre de copies de chromosomes moyenné au cours de l'induction et  $p(n_C)$  est la probabilité d'avoir  $n_C$  copies du chromosome en moyenne au cours de l'induction. En faisant de même pour  $\langle P_O \rangle$ , on obtient :

$$\langle P_V \rangle = \langle n_C \rangle \langle P_V^1 \rangle \quad (8.5)$$

$$\langle P_O \rangle = \langle n_P \rangle \langle P_V^1 \rangle \quad (8.6)$$

$$\text{d'où } \langle n \rangle = \frac{\langle n_P \rangle}{\langle n_C \rangle} = \frac{\langle P_O \rangle}{\langle P_V \rangle} \quad (8.7)$$

$\langle n \rangle$  désigne le PCN moyen par chromosome. Donc en mesurant  $\langle P_V \rangle$  et  $\langle P_O \rangle$ , nous avons bien accès au PCN moyen par chromosome.

De plus, notre induction dure une heure, soit environ deux cycles cellulaires à 37°C, et le nombre de copies du chromosome sur deux cycles entiers ne varie pas d'une bactérie à l'autre. Il semble raisonnable de poser  $n_C = \langle n_C \rangle = 1,7$  à 37°C [62]. Ainsi, nous avons même accès à  $\langle n_P \rangle = \frac{\langle P_O \rangle}{\langle P_V \rangle} \langle n_C \rangle$



**Remarque** :  $n_C$  et  $n_P$  ne sont pas entiers étant donné que le nombre de copies au cours de notre induction varie. Cependant, utiliser des intégrales plutôt que des sommes donne les mêmes conclusions. Nous garderons cette notation car elle est permet une compréhension plus "directe" de ce que nous faisons.

A l'aide de notre dispositif microfluidique, nous avons pu extraire les nombres de copies moyens pour les différents plasmides. Nous avons répété plusieurs fois chaque expérience. La dispersion entre plusieurs mesures similaires étaient typiquement un ordre de grandeur au dessus des incertitudes lié à la mesure (voir Annexes C et D). Ce sont ces dispersions qui nous ont servi à donner les incertitudes.

	miniF	miniR1par <sup>-</sup>	miniR1par <sup>+</sup>	miniColE1
$\langle P_V \rangle$	$27,1 \pm 0,8$	$28,5 \pm 0,8$	$26,5 \pm 0,4$	$25,7 \pm 1,9$
$\langle P_O \rangle$	$27,0 \pm 0,9$	$244 \pm 16$	$173 \pm 8$	$2167 \pm 89$
$\langle n_P \rangle = \langle n_C \rangle \langle P_O \rangle / \langle P_V \rangle$	$1,71 \pm 0,06$	$13,3 \pm 0,5$	$11,0 \pm 0,5$	$161 \pm 4$
$\langle n \rangle = \langle P_O \rangle / \langle P_V \rangle$	$1,0 \pm 0,04$	$7,8 \pm 0,2$	$6,5 \pm 0,3$	$95 \pm 2$
$[Bla] / [pyrF]$	0,5	3,2	3,8	23,4

**Le plasmide miniF** exhibe un nombre de copies moyen est de  $1,0 \pm 0,04$  par terminus du chromosome. Guyer et al. [64] donnent plutôt 1,4 copies par chromosome.

**Les plasmides miniR1par<sup>-</sup> et miniR1par<sup>+</sup>.** Leur PCN n'est jamais mentionné exactement dans la littérature. Cependant, il est admis que leur PCN moyen est compris entre 3 et 20 ce qui est cohérent avec les nombres que nous avançons. Par contre, nous avons mesuré une différence sensible entre les deux plasmides. Le PCN moyen du miniR1par<sup>-</sup> apparaît plus élevé que celui du miniR1par<sup>+</sup>. Nous pouvons donner deux explications à ce phénomène : 1- le nombre de copies moyen est réellement différent, 2- le rapporteur fluorescent sur le miniR1par<sup>+</sup> est rendu partiellement silencieux (voir 4.2.2). L'effet d'extinction de fluorescence peut-être très important (plus de 80% pour de fortes expression de *parB*). Cependant, ce niveau n'est peut-être jamais atteint dans notre système et/ou que transitoirement. En effet, Rodionov et al. [35], qui trouvent justement cet effet d'extinction, surexpriment ParB, la protéine responsable de l'effet, à l'aide d'un système d'expression exogène. Nous en discuterons plus loin (voir la partie 9.2).

**Le plasmide miniColE1** nous donne un PCN moyen plus haut que ce à quoi nous nous attendions. En effet, nos résultats de PCRs en temps réel nous donneraient plutôt un nombre en dessous d'un facteur 2. Cependant, comme précisé dans la partie 8.1, nous ne connaissons pas le rendement du kit d'extraction pour les plasmides et encore moins la relation entre la taille de l'ADN et le rendement de l'extraction.

## 8.5 Effet de la température sur les miniR1

Les miniR1 que nous avons utilisés ont un contrôle de leur réplication spécifiquement dépendant de la température. Lorsque la température baisse, leur fréquence d'initiation de réplication baisse ce qui aboutit à un nombre de copies moyen inférieur.

	miniR1par <sup>-</sup>	miniR1par <sup>+</sup>
$\langle n \rangle$ à 37°	$7,8 \pm 0,2$	$6,5 \pm 0,3$
$\langle n_P \rangle$ à 37°	$13,3 \pm 0,5$	$11,0 \pm 0,5$
$\langle n \rangle$ à 30°	$6,8 \pm 0,4$	$4,7 \pm 0,25$
$\langle n_P \rangle$ à 30°	$9,3 \pm 0,6$	$6,4 \pm 0,4$

Premièrement, nous observons bien une baisse des PCNs avec une baisse de température. De plus,  $\frac{\langle n_P(\text{miniR1par}^-) \rangle}{\langle n_P(\text{miniR1par}^+) \rangle}$  passe de 1,21 à 37°C et 1,45 à 30°C. Dit autrement, le rapport entre le PCN du miniR1par<sup>-</sup> et le PCN du miniR1par<sup>+</sup> augmente avec la baisse de température. Ceci suggère que l'effet du système de partitionnement se fait d'autant plus ressentir que le PCN est bas.

## Chapitre 9

# Variation du nombre de copies de plasmides

### 9.1 Correction de la fluorescence pour les variances

Afin d'obtenir les variances des  $P_V$  et  $P_O$ , nous ferons l'hypothèse que l'autofluorescence des bactéries n'est pas corrélée au niveau de fluorescence liée à l'expression de nos rapporteurs. Ainsi, nous avons obtenu :

$$\begin{aligned} \langle \Delta P_V^2 \rangle = \frac{1}{(a - \alpha\beta)^2} \{ & a^2 (\langle \Delta F_V^2 \rangle - \langle \Delta A_V^2 \rangle) \\ & + \alpha^2 \times (\langle \Delta F_O^2 \rangle - \langle \Delta A_O^2 \rangle) \\ & - 2\alpha a \times (\langle \Delta F_V \Delta F_O \rangle - \langle \Delta A_V \Delta A_O \rangle) \} \end{aligned} \quad (9.1)$$

$$\begin{aligned} \langle \Delta P_O^2 \rangle = \frac{1}{(a - \alpha\beta)^2} \{ & (\langle \Delta F_O^2 \rangle - \langle \Delta A_O^2 \rangle) \\ & + \beta^2 \times (\langle \Delta F_V^2 \rangle - \langle \Delta A_V^2 \rangle) \\ & - 2\beta \times (\langle \Delta F_O \Delta F_V \rangle - \langle \Delta A_O \Delta A_V \rangle) \} \end{aligned} \quad (9.2)$$

$$\langle \Delta P_V \Delta P_O \rangle = \frac{1}{(a - \alpha\beta)} \{ \langle \Delta F_O \Delta F_V \rangle - \langle \Delta A_O \Delta A_V \rangle - a\alpha \langle \Delta P_O^2 \rangle - \beta \langle \Delta P_V^2 \rangle \} \quad (9.3)$$

avec  $\Delta P_V = P_V - \langle P_V \rangle$ ,  $\Delta P_O = P_O - \langle P_O \rangle$ ,  $\Delta F_V = F_V - \langle F_V \rangle$ ,

$\Delta F_O = F_O - \langle F_O \rangle$ ,  $\Delta A_V = A_V - \langle A_V \rangle$  et  $\Delta A_O = A_O - \langle A_O \rangle$

## 9.2 Bruit d'expression lié au plasmide

Une fois que nous avons vérifié que nous pouvions estimer le PCN moyen pour chacun des nos plasmides, nous avons extrait les bruits d'expression de chacun des rapporteurs. Pour cela, il suffit de calculer :

$$\eta_V = \frac{\sqrt{\langle (P_V - \langle P_V \rangle)^2 \rangle}}{\langle P_V \rangle}$$

$$\eta_O = \frac{\sqrt{\langle (P_O - \langle P_O \rangle)^2 \rangle}}{\langle P_O \rangle}$$

De la même manière que dans 8.4, les incertitudes sont calculées à partir de la dispersion entre les mesures effectuées sur les mêmes souches.

	miniF	miniR1par <sup>-</sup>	miniR1par <sup>+</sup>	miniColE1
$\eta_V (\times 10^{-2})$	$44,3 \pm 0,6$	$43,5 \pm 0,8$	$45 \pm 3$	$45,2 \pm 1,2$
$\eta_O (\times 10^{-2})$	$66,6 \pm 1,8$	$54,4 \pm 1,2$	$49,8 \pm 1,5$	$46,2 \pm 0,9$

Les mesures que nous avons faites à 37°C montrent que, grossièrement, le bruit d'expression est sensiblement plus fort pour les plasmides à bas PCN que pour les plasmide à haut PCN.

On remarque cependant que  $\eta_O$  est plus petit pour le miniR1par<sup>+</sup> que pour le miniR1par<sup>-</sup> (voir figure 9.1). Nous avons conclu dans 8.4 que  $\langle n(\text{miniR1par}^+) \rangle$  était plus faible que  $\langle n(\text{miniR1par}^-) \rangle$  pour deux raisons possibles, soit à cause d'un PCN moyen plus faible, soit à cause de l'effet d'extinction. L'extinction advient du fait de la diminution du taux de transcription du gène ce qui se traduit non seulement par une baisse de production de la protéine mais aussi par une plus grande variabilité [9]. Or, nous observons plutôt une baisse de la variabilité. On peut donc en déduire que le PCN du miniR1par<sup>+</sup> est plus petit que le PCN du miniR1par<sup>-</sup>.

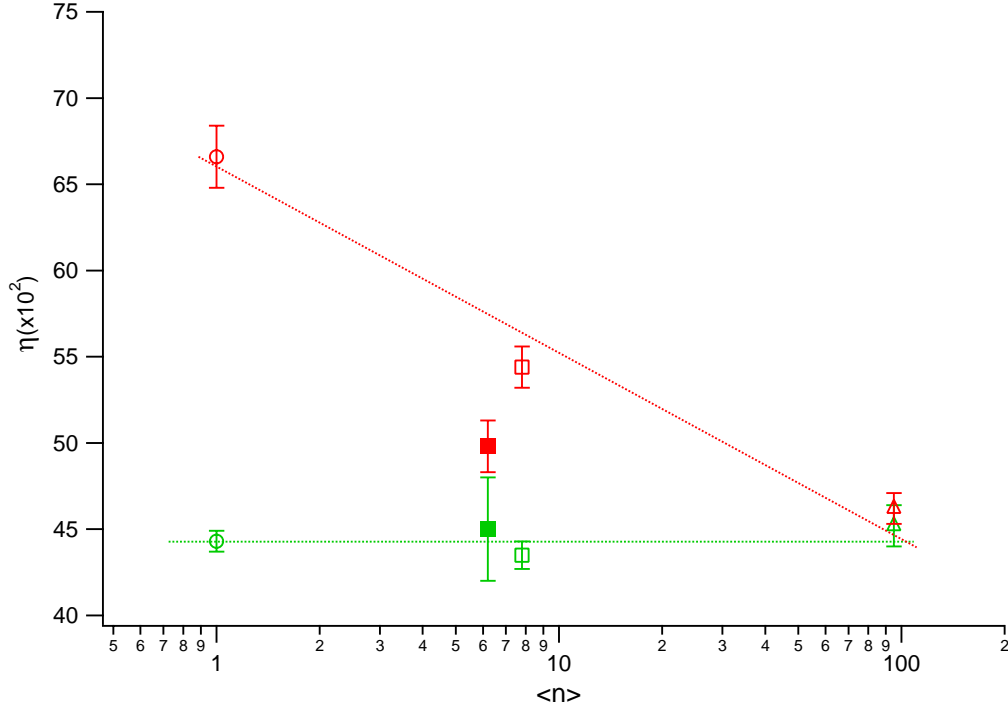


FIG. 9.1: Sont représentés les bruits d'expression  $\eta_V$  en vert et  $\eta_O$  en rouge en fonction du PCN moyen  $\langle n \rangle$  pour tous les plasmides que nous avons étudié. Les différents symboles : (○) pour le miniF, (□) pour le miniR1par<sup>-</sup>, (■) pour le miniR1par<sup>-</sup> et finalement (Δ) pour le miniColE1. L'échelle des  $\langle n \rangle$  est logarithmique. Les droites vertes et rouges ne sont que des guides optiques. Les barres d'erreurs horizontales n'ont pas été rajoutées car elles sont environ de la taille des marqueurs.

### 9.3 Variance du nombre de copies de plasmides

En partant des équations 8.3 et 8.4,

$$P_V^2 = \sum_{i=1}^{n_C} \sum_{j=1}^{n_C} P_V^i P_V^j \quad (9.4)$$

$$P_O^2 = \sum_{i=1}^{n_P} \sum_{j=1}^{n_P} P_O^i P_O^j \quad (9.5)$$

$$P_O P_V = \sum_{i=1}^{n_C} \sum_{j=1}^{n_P} P_V^i P_O^j \quad (9.6)$$

On pose ensuite les hypothèses suivantes :

- en moyenne, les corrélations entre l'expression d'un des rapporteurs à partir d'une seule copie et celle de l'autre rapporteur à partir d'une autre copie est la même.

Soit :

$$\langle P_O^i P_V^j \rangle = \langle P_O^1 P_V^1 \rangle \quad \forall i, j$$

- en moyenne, les corrélations entre l'expression d'un des rapporteurs à partir d'une seule copie et celle du même rapporteur à partir d'une autre copie est toujours la même. Soit :

$$\langle P_O^i P_O^j \rangle = \langle P_V^i P_V^j \rangle = \langle P_O^1 P_V^1 \rangle \quad \forall i \neq j.$$

- en moyenne, l'autocorrélation de l'expression d'un rapporteur à partir d'une seule copie est la même quelque soient le rapporteur et la copie considérés. Soit :

$$\langle P_O^i P_O^i \rangle = \langle P_V^i P_V^i \rangle = \langle (P_V^1)^2 \rangle \quad \forall i$$

Lorsqu'on écrit  $\langle P_V^2 \rangle$  et  $\langle P_O^2 \rangle$  et  $\langle P_O P_V \rangle$ , nous faisons apparaître  $n_C$  et  $n_P$ , leurs probabilités respectives  $p(n_C)$  et  $p(n_P)$  et leur probabilité conjointe  $p(n_C, n_P)$  :

$$\langle P_V^2 \rangle = \sum_{n_C} p(n_C) \left\{ \sum_{i=1}^{n_C} \langle (P_V^1)^2 \rangle + \sum_{i=1, j=1, i \neq j}^{n_C} \langle P_O^1 P_V^1 \rangle \right\} \quad (9.7)$$

$$\langle P_O^2 \rangle = \sum_{n_P} p(n_P) \left\{ \sum_{i=1}^{n_P} \langle (P_V^1)^2 \rangle + \sum_{i=1, j=1, i \neq j}^{n_P} \langle P_O^1 P_V^1 \rangle \right\} \quad (9.8)$$

$$\langle P_O P_V \rangle = \sum_{n_C} \sum_{n_P} p(n_C, n_P) \left\{ \sum_{n_C} \sum_{n_P} \langle P_O^1 P_V^1 \rangle \right\} \quad (9.9)$$

Nous avons déjà vu que  $n_C = \langle n_C \rangle = 1,7$  à  $37^\circ\text{C}$  (voir la partie 8.4). On en déduit que  $\langle (n_C)^2 \rangle = \langle n_C \rangle^2$  et que  $\langle P_O^1 P_V^1 \rangle = \frac{\langle P_O P_V \rangle}{\langle n_C \rangle \langle n_P \rangle}$ .

A partir des équations 9.7, 9.8 et 9.9, on peut écrire  $\langle (n_P)^2 \rangle$  en fonction des valeurs que nous mesurons ou connaissons :

$$\langle (n_P)^2 \rangle = \langle n_P \rangle \left\{ 1 + \frac{\langle n_C \rangle}{\langle P_O P_V \rangle} \left[ \langle P_O^2 \rangle - \frac{\langle n_P \rangle}{\langle n_C \rangle} \langle P_V^2 \rangle \right] \right\} + \frac{\langle n_P \rangle}{\langle n_C \rangle} \{ \langle n_C \rangle^2 - \langle n_C \rangle \} \quad (9.10)$$

Nous pouvons désormais calculer  $\langle (n_P)^2 \rangle$  et  $\langle n_P \rangle$ . Nous en déduisons alors

$$\begin{aligned} \sigma_{n_P} &= \sqrt{\langle (n_P)^2 \rangle - \langle (n_P) \rangle^2} \\ \sigma_n &= \frac{\sigma_{n_P}}{\langle n_C \rangle} \\ \eta_n &= \frac{\sqrt{\langle (n_P)^2 \rangle - \langle (n_P) \rangle^2}}{\langle (n_P) \rangle} = \frac{\sqrt{\langle (n)^2 \rangle - \langle (n) \rangle^2}}{\langle (n) \rangle} \end{aligned}$$

	miniF	miniR1par <sup>-</sup>	miniR1par <sup>+</sup>	miniColE1
$\langle n \rangle = \langle P_O \rangle / \langle P_V \rangle$	$1,0 \pm 0,04$	$7,8 \pm 0,2$	$6,5 \pm 0,3$	$95 \pm 2$
$\sigma_n$	$0,50 \pm 0,02$	$2,7 \pm 0,12$	$1,9 \pm 0,13$	$24 \pm 2$
$\langle n_P \rangle$	$1,71 \pm 0,06$	$13,3 \pm 0,5$	$11,0 \pm 0,5$	$161 \pm 4$
$\sigma_{n_P}$	$0,70 \pm 0,03$	$4,2 \pm 0,2$	$3,1 \pm 0,3$	$40 \pm 4$
$\eta_n (\times 10^{-2})$	$46 \pm 2$	$34 \pm 2$	$29,2 \pm 0,6$	$25 \pm 3$

Avec les hypothèses que nous avons posé pour calculer les  $\sigma_{n_P}$  et les  $\eta_n$ , la variabilité du nombre de chromosome moyenné au cours de l'induction par bactérie  $\eta_{n_C} = \frac{\sqrt{\langle (n_C)^2 \rangle - \langle (n_C) \rangle^2}}{\langle (n_C) \rangle}$  est nul. Nous voyons donc que tous les plasmides ont une variabilité de leur nombre de copies significativement plus grande que celle du chromosome.

**Le plasmide miniF** est le plasmide qui a le plus fort  $\eta_n$ . Il est important de comprendre cela comme un effet "mécanique" de son bas PCN.

**Les plasmides miniR1par<sup>-</sup> et miniR1par<sup>+</sup>.** Les mesures nous confirment que le système de partitionnement n'a pas seulement pour effet de répartir au moins une copie au sein de chacune des cellules filles mais qu'il participe activement à une répartition plus équilibrée des multiples copies du plasmides entre les bactéries filles.

**Le plasmide miniColE1** exhibe un très grand écart-type en PCN et la plus petite variabilité. Cela est la conséquence de son grand PCN.

Nous avons désormais les deux premiers moments de la distribution des PCN pour nos plasmides sans pour autant avoir la forme de la distribution. Paulsson et Ehrenberg [44] lors de leur développement sur le système de régulation de la réplication des plasmides de type R1 montrent que dans le cas d'un plasmide R1 sans système de partition la distribution de  $n_P$  devrait être proche d'une poissonnienne<sup>1</sup>. Nous pouvons déjà tester si la distribution peut être poissonnienne (voir figure 9.2). En effet, pour une telle distribution, la variance est égale à la moyenne. Les distributions du miniF et du miniColE1 sont respectivement beaucoup plus étroite et beaucoup plus large qu'une distribution poissonnienne.

<sup>1</sup>Cependant, leur développement fait intervenir la distribution binomiale des plasmides entre les bactéries filles. Ceci ne semble donc pas compatible avec un plasmide R1 portant un système de partition.

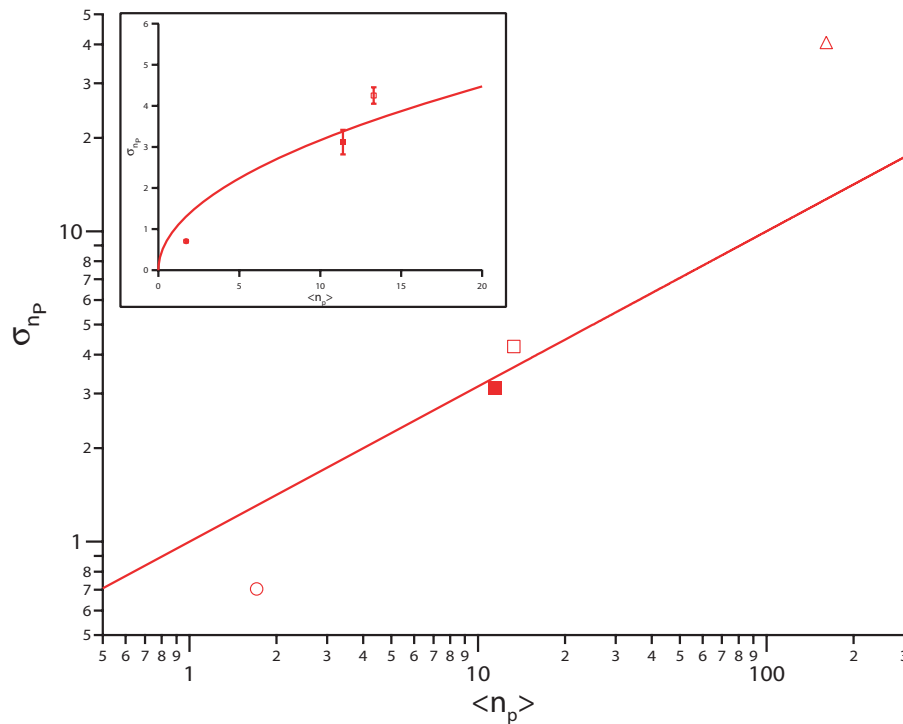


FIG. 9.2: Sont représentés les écarts-types des PCN en fonction des PCN moyens pour les différents plasmides. Les traces continues correspondent au cas d'une distribution poissonnienne. Les différents symboles : ( $\circ$ ) pour le miniF, ( $\square$ ) pour le miniR1par<sup>-</sup>, ( $\blacksquare$ ) pour le miniR1par<sup>-</sup> et finalement ( $\triangle$ ) pour le miniColE1. L'insert correspond à un agrandissement de la même représentation avec des échelles linéaires.

En fait, seul les miniR1 avec et sans système de partition s'en rapprochent.

## 9.4 Une méthode "graphique"

Nous avons essayé d'obtenir la distribution du PCN par une méthode alternative, peut-être plus intuitive.

### 9.4.1 Discerner entre plusieurs copies d'un gène

Dans la partie 8.4, nous avons comparé  $\langle P_O \rangle$  et  $\langle P_V \rangle$  afin d'estimer le PCN moyen par bactérie. Nous avons alors vérifié que  $\langle P_O \rangle / \langle P_V \rangle$  était bien linéaire avec le nombre de copies du gène rapporteur au sein de la bactérie. On peut alors se demander si  $P_O / P_V$  ne nous donnerait pas accès au PCN dans chacune des bactéries. Pour cela, nous devons



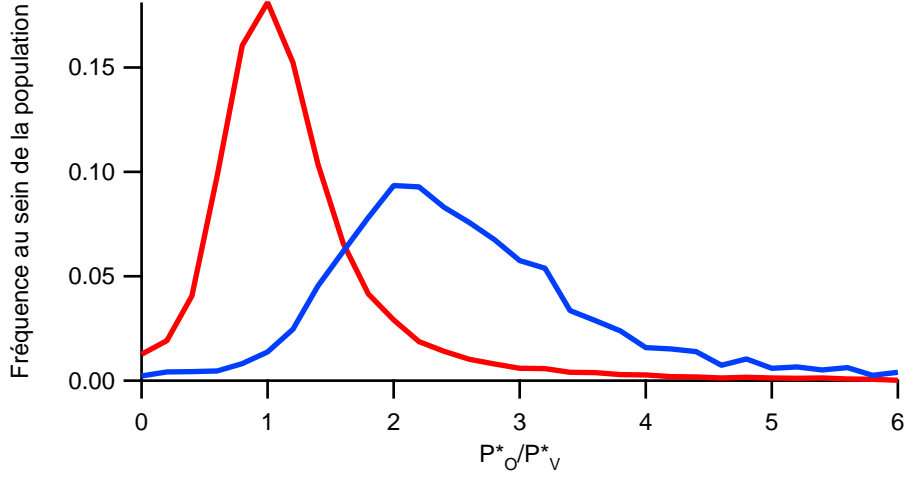


FIG. 9.3: Distributions de  $P_O^*/P_V^*$  pour le miniF contenant une simple construction (en rouge) et pour le miniF contenant un dimère de la construction (en bleu).

corriger les mesures de fluorescences pour l'autofluorescence et pour les fuites de détection d'un canal sur l'autre au sein de chaque bactérie. Or cela n'est pas possible exactement. Cependant, nous pouvons corriger pour une autofluorescence moyenne puis pour les fuites ce qui nous donne :

$$P_V^* = \frac{a}{a - \alpha\beta} \{ (F_V - \langle A_V \rangle) - \frac{\alpha}{a} (F_O - \langle A_O \rangle) \} \quad (9.11)$$

$$P_O^* = \frac{1}{a - \alpha\beta} \{ (F_O - \langle A_O \rangle) - \beta (F_V - \langle A_V \rangle) \} \quad (9.12)$$

Nous avons à notre disposition le plasmide miniF contenant une fois notre construction (mOrange) mais aussi le plasmide miniF contenant un dimère de notre construction. Lorsqu'on trace les distributions de  $P_O^*/P_V^*$  pour ces deux souches, on observe un recouvrement non négligeable (voir figure 9.3). Même si les deux distributions semblent discernables, à l'aide d'une mesure, les variations de PCN entre bactéries ne seront pas aussi discrètes. Comme nous induisons la production de la mOrange pendant une heure, le PCN que l'on mesure sur une bactérie est en fait déjà la moyenne de son PCN au cours de son histoire durant l'heure précédente.

Par cette opération, nous pensons éliminer les variations qui influencent de la même

manière  $P_O$  et  $P_V$  <sup>2</sup>.

Restent alors la variation du PCN et la variation d'expression. Nous ne savons pas prédire cette dernière. On pourra donc seulement dire que les variations de  $P_O^*/P_V^*$  surestiment la variation du PCN.

On pourra néanmoins utiliser la distribution des  $P_O^*/P_V^*$  afin de sonder la distribution des  $n_P$ . Considérer que cela est faisable revient en fait à considérer que les bactéries qui ont le même rapport de fluorescence ont à peu près le même PCN.

### 9.4.2 Construction de la distribution

Nous noterons désormais  $P_O^*/P_V^* = div$ . Ainsi,  $\langle div \rangle$ ,  $\sigma_{div}$  et  $\eta_{div}$  seront la moyenne, l'écart-type et le bruit de  $div$  <sup>3</sup>

	miniF	miniR1par <sup>-</sup>	miniR1par <sup>+</sup>	miniColE1
$\langle n \rangle = \langle P_O \rangle / \langle P_V \rangle$	$1,0 \pm 0,04$	$7,8 \pm 0,2$	$6,2 \pm 0,3$	$95 \pm 2$
$\sigma_n$	$0,50 \pm 0,02$	$2,7 \pm 0,12$	$1,9 \pm 0,13$	$24 \pm 2$
$\eta_n (\times 10^{-2})$	$46 \pm 2$	$34 \pm 2$	$29,1 \pm 0,6$	$25 \pm 3$
$\langle div \rangle$	$1,01 \pm 0,03$	$8,0 \pm 0,4$	$6,7 \pm 0,2$	$98 \pm 2$
$\sigma_{div}$	$0,52 \pm 0,02$	$3,2 \pm 0,2$	$2,8 \pm 0,2$	$36 \pm 2$
$\eta_{div} (\times 10^{-2})$	$51 \pm 1$	$40 \pm 2$	$41 \pm 1$	$37 \pm 3$

La moyenne de  $div$  donne des valeurs comparables à ce que nous avons obtenu pour  $\langle n \rangle$ . Cela est en partie compréhensible car en prenant les moyennes des équations 9.11 et 9.11, on obtient :  $\langle P_V^* \rangle = \langle P_V \rangle$  et  $\langle P_O^* \rangle = \langle P_O \rangle$ . Ceci implique donc que  $\langle \frac{P_O^*}{P_V^*} \rangle \simeq \frac{\langle P_O^* \rangle}{\langle P_V^* \rangle}$  ce qui n'est pas évident à montrer.

Ensuite, concernant  $\sigma_{div}$  et  $\eta_{div}$ , nous trouvons des valeurs significativement plus élevées que  $\sigma_n$  et  $\eta_n$ . De plus, les deux miniR1 ont le même  $\eta_{div}$ . Il semble donc que nous surestimons sensiblement la variabilité du nombre du PCN avec  $div$  et que cette surestimation n'est

<sup>2</sup>Nous n'utiliserons pas le terme de variabilité "extrinsèque". Nous ne pouvons pas faire la même distinction entre variabilité intrinsèque et extrinsèque que dans Swain et al. [65]. En effet, si on pose le bruit intrinsèque et le bruit extrinsèque comme indépendants, on se rend compte que leur définition par rapport aux fluorescences mesurées change en fonction du nombre de plasmides ce qui rend leur interprétation plus ambiguë.

<sup>3</sup> $\sigma_{div} = \sqrt{\langle div^2 \rangle - \langle div \rangle^2}$  et  $\eta_{div} = \sqrt{\frac{\langle div^2 \rangle - \langle div \rangle^2}{\langle div \rangle^2}}$

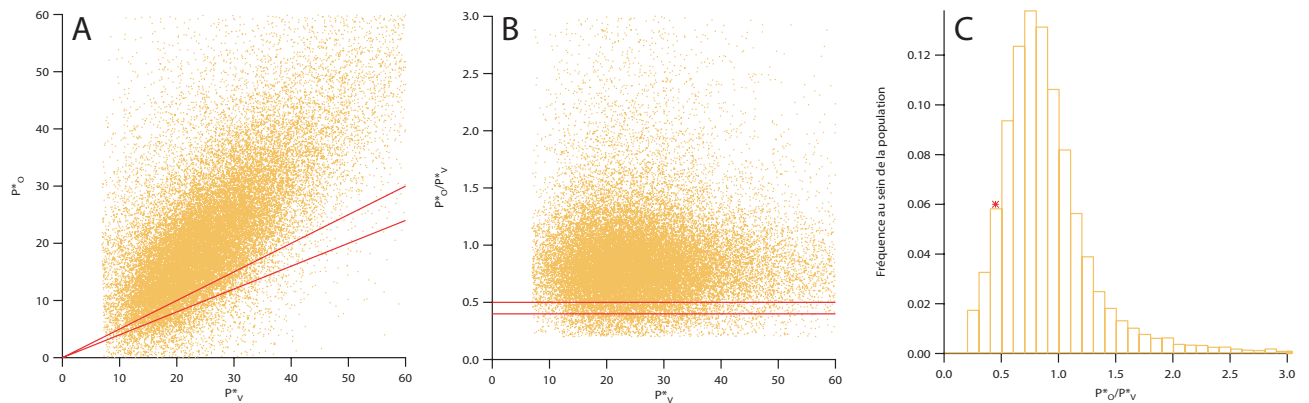


FIG. 9.4: Construction de l'histogramme de  $P_O^*/P_V^*$  pour une population contenant le miniF. (A) Diagramme de dispersion des bactéries. On représente  $P_O^*$  par rapport à  $P_V^*$  toute la population. Chaque point correspond à une bactérie. (B) Diagramme de dispersion des mêmes bactéries. On représente cette fois  $P_O^*/P_V^*$  par rapport à  $P_V^*$  toute la population. Chaque point correspond à une bactérie. (C) Histogramme normalisé des  $P_O^*/P_V^*$ . Les droites rouges "encadrent" les bactéries qui auraient entre 0,4 et 0,5 PCN par chromosome. Le bin marqué d'une étoile est formé à partir de ces bactéries.

pas la même pour tous les plasmides. La tendance générale de la variabilité est quand même respectée

Clairement, *div* permet une bonne estimation du PCN moyen par chromosome. Par contre, son utilisation pour l'estimation de la variabilité est à faire avec précaution. Nous avons alors construit les histogrammes des *div* (voir figure 9.4).

Quel que soit le plasmide, l'écart entre la valeur la plus probable de *div* et sa moyenne est important (voir figure 9.5). Cela est dû au fait que les distributions que l'on obtient ne sont pas du tout symétriques.

De plus, on différencie assez mal les distributions des *div* normés pour chacun des plasmides. Ainsi, si nous voulons déterminer la forme de la distribution des PCNs avec les mêmes données, il sera nécessaire d'explorer d'autres méthodes de regroupement plus complexes que de simplement considérer le rapport des fluorescences des deux rapporteurs.

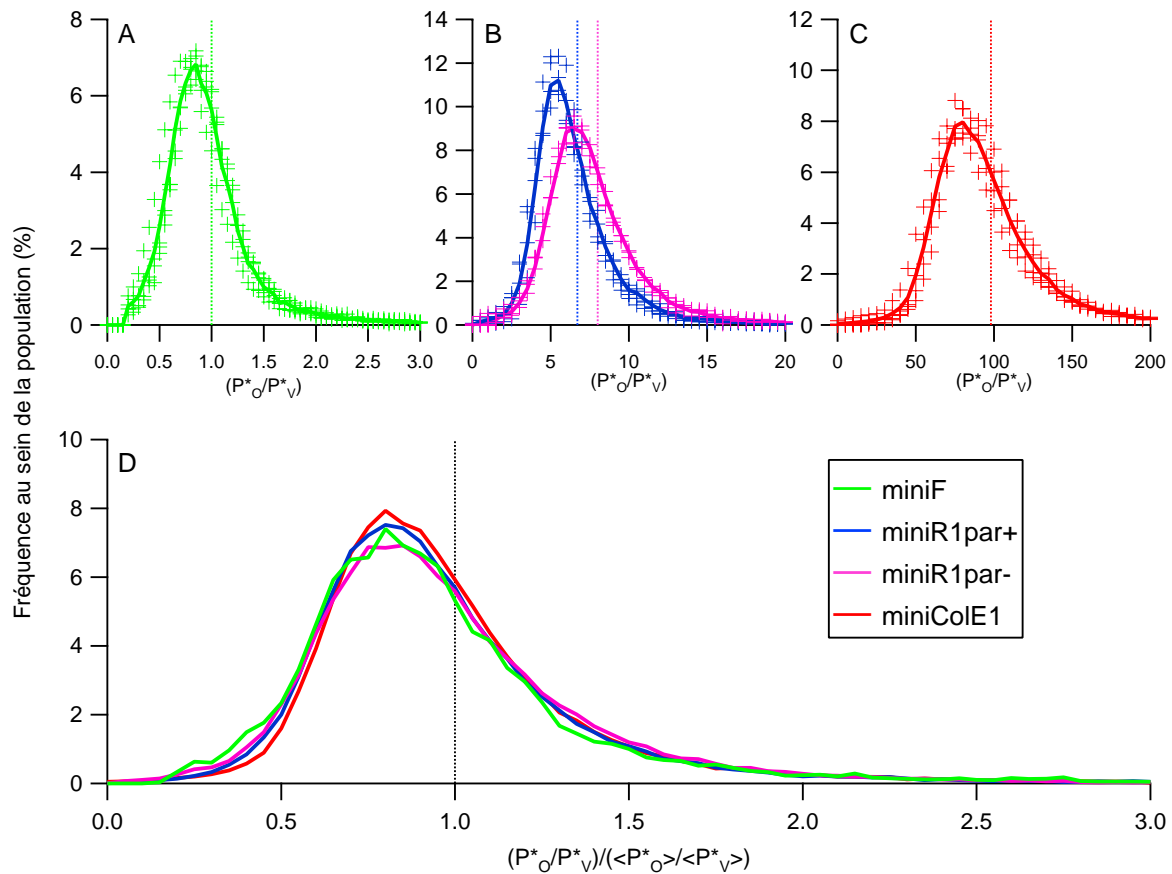


FIG. 9.5: (A-C) Histogrammes de  $P^*_O/P^*_V$  pour le miniF (vert) , le miniR1par<sup>+</sup> (bleu) , le miniR1par<sup>-</sup> (violet) et le miniColE1 (rouge). Pour chaque plasmide, les histogrammes pour plusieurs expériences sont représentés par des marqueurs (+) et la moyenne de ces histogrammes par un trait continu.(D) Histogrammes moyens sur plusieurs expériences de  $\frac{P^*_O/\langle P^*_O \rangle}{P^*_V/\langle P^*_V \rangle}$  pour chaque plasmide. Les lignes verticales indiquent les valeurs moyennes pour chacun des plasmides.

## Chapitre 10

# Changement de concentration d'antibiotique

Nous avons exclusivement utilisé l'ampicilline comme antibiotique. Cet antibiotique fait partie de la grande famille des pénicillines. La résistance à cet antibiotique est codée par le gène *bla*. Ce gène produit une enzyme qui dégrade l'ampicilline.

### 10.1 Qualité de la résistance avec le nombre de copies du plasmide

Nous avons voulu savoir si la résistance portée par nos plasmides permettait aux bactéries de croître en milieu fortement concentré en antibiotique (voir figure [10.1](#)). Pour cela, nous avons étalé nos différentes souches sur une boîte de pétri. Ensuite, nous avons appliqué au centre de la boîte un morceau de papier (patch) imbibé l'ampicilline à très forte concentration (1000× la concentration recommandée pour le maintien d'un plasmide contenant la résistance à l'antibiotique). Les souches ne contenant pas le gène de résistance n'ont pu croître que loin du patch. Pour les souches contenant les plasmides, seul le miniF n'a pas pu croître aux abords du "patch".

Ainsi, le nombre de copies du gène de résistance à l'ampicilline ne semble pas être important pour la qualité de la résistance tant que nous nous plaçons à un PCN supérieur à celui du miniF. Nous avons néanmoins voulu savoir si des effets plus fins n'étaient pas décelables à l'aide de notre système de mesure.

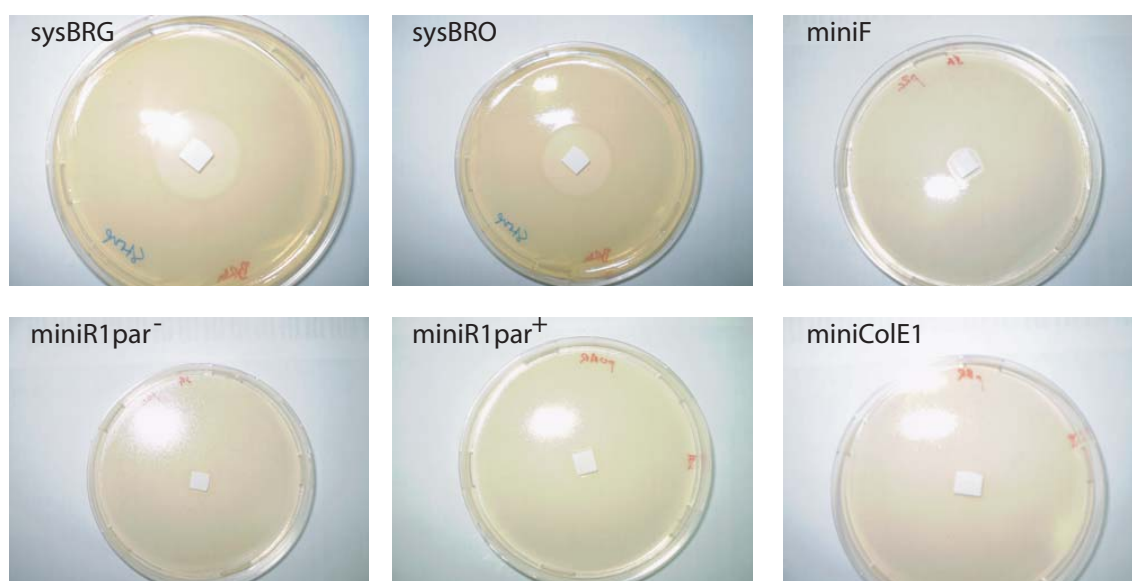


FIG. 10.1: Une solution saturée de bactérie est diluée 30 fois. 2ml de cette solution sont étalés sur la boîte de pétri. Au bout de 15 minutes, le surnageant est retiré et les boîtes sont laissées à température ambiante pendant une dizaine de minutes. On rajoute ensuite un morceau de papier d'environ 5mm de côté imbibé de 10 $\mu$ l d'ampicilline à 50mg par ml. Les souches SysBRG et SysBRO ne contiennent pas le gène de résistance à l'ampicilline. Les autres souches contiennent la résistance à l'ampicilline sur le plasmide.

## 10.2 Constance du PCN et de sa variabilité

Pour chacune des souches que nous avons à disposition, nous sommes partis d'une solution de bactéries fraîchement saturés.

Ensuite, par des dilutions successives, nous avons fait croître les différentes souches sur 90 générations dans un milieu nutritif contenant 0, 50 ou 250 $\mu$ g par ml.

Parmi les souches ayant cru dans un milieu sans antibiotiques, seule celle contenant initialement le plasmide miniR1par<sup>-</sup> exhibait une sous population ayant perdu le plasmide. Nous en parlerons dans la partie suivante (10.3).

Dans cette partie, nous ne traiterons que des populations ou sous populations ayant conservé le plasmide. Nous n'avons fait ces mesures qu'une seule fois. C'est pour cela que nous n'avons pas mis d'incertitudes sur les valeurs mentionnées. Néanmoins, les valeurs ne sont jamais très éloignées des valeurs obtenues pour la culture de départ. Nous n'avons

donc trouvé de changement significatif ni dans  $\langle n \rangle$  ni dans  $\sigma_n$ .

	miniF	miniR1par <sup>-</sup>	miniR1par <sup>+</sup>	miniColE1
$\langle n \rangle$ initial	$1,0 \pm 0,04$	$7,8 \pm 0,2$	$6,5 \pm 0,3$	$95 \pm 2$
$\sigma_n$ initial	$0,50 \pm 0,02$	$2,7 \pm 0,12$	$1,9 \pm 0,13$	$24 \pm 2$
$\langle n \rangle$ avec $0\mu\text{g}$ par ml	0,99	7,6	6,9	91
$\sigma_n$ avec $0\mu\text{g}$ par ml	0,43	2,48	2,1	20
$\langle n \rangle$ avec $50\mu\text{g}$ par ml	0,99	7,79	6,3	97,2
$\sigma_n$ avec $50\mu\text{g}$ par ml	0,39	$2,5 \pm 0,2$	1,8	20
$\langle n \rangle$ avec $250\mu\text{g}$ par ml	1,08	8,1	6,49	94
$\sigma_n$ avec $250\mu\text{g}$ par ml	0,39	$2,8 \pm 0,2$	1,9	23

### 10.3 Perte du plasmide

Nous n'avons pu observer une perte de plasmide que sur les bactéries contenant initialement le plasmide miniR1par<sup>-</sup>. La fraction de la population ayant conservé le plasmide  $\Psi^+$  dépend du nombre de génération passé dans le milieu sans ampicilline. Ainsi, au fur et à mesure des générations, une sous population grandissante de bactéries se retrouve sans plasmide.

Nous avons aussi estimé  $\Psi^+$  en isolant un certain nombre de clones puis en testant chacune d'elles pour la résistance à l'ampicilline. En pratique, il suffit d'étaler une solution de bactéries sur une boîte de pétri afin d'obtenir des colonies isolées. Ensuite, on réplique<sup>1</sup> cette boîte sur une autre boîte de pétri supplémentée d'ampicilline. C'est la méthode de "replica plating". La proportion de colonies ayant cru sur les deux boîtes nous donne  $\Psi^+$ . De plus, à l'aide des données de fluorescence de chaque bactérie obtenu avec notre système de détection, on peut facilement différencier les bactéries présentant une faible fluorescence rouge du reste de la population (voir la figure 10.2). On peut raisonnablement identifier  $\Psi^+$  avec la proportion de bactérie exhibant une forte fluorescence. Pour les populations que nous avons à considérer, une vingtaine de bactéries seulement sur plusieurs dizaines de milliers ne pouvaient être classés sans ambiguïté.

<sup>1</sup>Pour répliquer une boîte de pétri sur une autre, il suffit de mettre au contact un feutre avec la première boîte puis mettre le feutre au contact de la deuxième boîte.

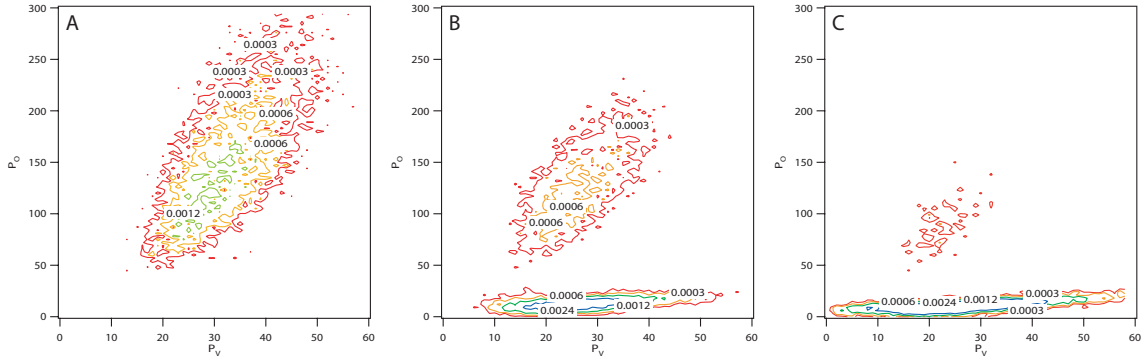


FIG. 10.2: Nous avons d'abord construit un histogramme normé à deux dimensions du diagramme de dispersion des bactéries par rapport à  $P_V$  et  $P_O$ . Sont ensuite représentés les mêmes iso-densités de cet histogramme pour la souche contenant le  $\text{miniR1par}^-$  : (A) qui évolué dans un milieu contenant  $50\mu\text{g}$  par ml d'ampicilline, (B) qui a évolué pendant 54 générations dans un milieu sans ampicilline et (C) qui a évolué pendant 99 génération dans un milieu sans ampicilline.

	Estimation par "replica plating"	Estimation par regroupement des données de fluorescence
$\Psi^+(54)$	$50 \pm 5$	$60,3 \pm 0,3$
$\Psi^+(99)$	$12 \pm 3$	$16,4 \pm 0,6$

A partir de  $\Psi^+$ , il est possible d'extraire un taux de perte de plasmide  $\theta$ . Ainsi Boe et Rasmussen [66] [67] proposent et vérifient deux méthodes pour estimer  $\theta$ .

La première n'est valable que lorsque le temps de division des bactéries portant le plasmide  $T^+$  et le temps de division des bactéries n'en portant plus  $T^-$  sont égaux. Dans ce cas, seul le taux de perte détermine  $\Psi^+$ . On a alors :

$$\theta_1 = 2 \left\{ 1 - \left( \frac{\Psi^+(g)}{\Psi^+(0)} \right)^{1/g} \right\}$$

où  $\Psi^+(g)$  est la fraction de la population portant le plasmide après  $g$  génération dans le milieu sans antibiotique.

Ici,  $\theta_1$  ne dépend pas de la génération à laquelle l'estimation est faite. Or, nous avons estimé  $\theta_1(54) \simeq 1,9\%$  et  $\theta_1(99) \simeq 3,6\%$ . Nous en déduisons que nous ne vérifions pas l'hypothèse de départ.



La deuxième méthode proposée nécessite de connaître  $\Psi^+$  pour au moins deux générations. On détermine d'abord  $\gamma = T^+/T^-$  avec

$$\ln \left( \frac{\Psi^-(g)}{\Psi^+(g)} \right) = \ln \left( \frac{\Psi^-(0)}{\Psi^+(0)} \right) + (\gamma - 1)g \ln 2 \quad (10.1)$$

Nous obtenons alors  $\gamma \simeq 1,066$ . Nous avons mesuré des temps de division (voir la partie 7.1) de l'ordre de 30 minutes. Un tel  $\gamma$  correspondrait à une différence de deux minutes, ce que nous ne pouvons observer étant donné la faible précision des données. Une fois  $\gamma$  déterminé, nous utilisons

$$\theta_2 \approx 2 \frac{\Psi^-}{\Psi^+} (1 - 2^{\gamma-1}) (1 - 2^{g(\gamma-1)}) \quad (10.2)$$

Nous obtenons alors  $\theta_2(54) \approx 0,57\%$  et  $\theta_2(99) \approx 0,53\%$ . Nous trouvons donc un bon accord entre nos deux mesures.

Un deuxième fait vient confirmer que nous avons bien estimé le taux de perte. Paulsson et Ehrenberg [44] ont calculé les taux de perte pour un système de régulation de type R1 pour plusieurs PCN moyen et le taux de perte que nous estimons correspond bien au PCN que nous avons mesuré.

Nous concluons que nous avons une estimation du taux de perte cohérente avec elle-même et avec la littérature. Cependant, afin d'en être sûr, il faudrait faire d'autres mesures à d'autres générations et à d'autres températures.

## Quatrième partie

### Conclusions

Nous voulions mesurer la variation du nombre de copies de plasmides au sein de populations monoclonales de bactéries.

Pour cela, nous avons construit des souches bactériennes exprimant deux rapporteurs fluorescents, l'un (eGFP) à partir du chromosome et l'autre (mOrange) à partir d'un des plasmides qui nous intéressent.

De plus, nous avons développé un système microfluidique et optique permettant la mesure de bas niveau de fluorescence sur des bactéries isolées.

Nous avons vérifié que nous obtenions bien des PCN moyens cohérents à la fois avec la littérature et avec les mesures "en volume" que peuvent nous donner les instruments de la biologie moléculaire.

Après avoir posé certaines hypothèses que nous estimons très raisonnables, nous avons pu extraire de nos mesures à la fois le PCN moyen de la population  $\langle n_P \rangle$  mais aussi l'écart-type du PCN  $\sigma_{n_P}$  et la variabilité du PCN  $\eta_n$  au sein de la population. Nous pensons que  $div = P_O^*/P_V^*$  peut nous donner une représentation visuelle raisonnable de la distribution de PCN dans une population.

A l'aide de l'ensemble de ces mesures, nous avons ainsi pu :

1. vérifier que le mini-F régule en valeur absolue très bien son PCN.
2. montrer que le système de partition participe aide à la répartition de l'ensemble des plasmides entre les bactéries filles.
3. mettre en évidence que l'effet du système de partition dans un miniR1 n'est pas seulement de resserrer la distribution du PCN au sein de la population mais aussi de diminuer le nombre de copies du plasmide.
4. voir que le miniColE1 dont on considère généralement qu'il a une régulation relâchée est le plasmide présentant le plus grand  $\sigma_{n_P}$  mais aussi le plus petit des  $\eta_n$  des trois systèmes que nous avons étudié.

Enfin, nous avons essayé de changer les caractéristiques de chacun des plasmides en faisant croître les bactéries dans un milieu avec une concentration d'antibiotiques différent. Cependant, nous n'avons observé de changement ni dans le PCN moyen ni dans  $\sigma_{n_P}$  sur une centaine de génération. Par contre, nous avons pu observer la perte de plasmide lorsque nous faisons croître des bactéries contenant le miniR1par<sup>-</sup> dans un milieu sans antibiotiques. Nous avons ainsi estimé le taux de perte de ce plasmide à 37°C et la diffé-

rence de temps de division entre une bactérie portant le *miniR1par<sup>-</sup>* et une autre l'ayant perdu.

Maintenant que tous nos plasmides sont bien caractérisés, plusieurs directions sont envisageables. Par exemple, nous pourrions étudier l'effet de gènes tels que *ccdA* et *ccdB* [68] sur le PCN et sur sa variabilité. Ces deux gènes forcent l'addiction de la bactérie au plasmide. On pourrait aussi étudier l'effet de mutations tels que *pcnB* supposé modifié le PCN des plasmides ColE1 [69].

Cependant, une direction retiendra notre attention. Nous voulons faire évoluer le PCN au sein de la population à l'aide de changement "environnementaux" et suivre sa dynamique. Il a déjà été montré que cela était possible [70], mais dans notre cas, la résistance portée par nos plasmides (ampicilline) n'est pas forcément la plus judicieuse et l'influence du PCN sur le métabolisme ne semble pas suffisant.

Nous pourrions changer la résistance utilisée. Il existe certaines résistances comme CAT pour le chloremphénicol qui plus elles sont exprimées plus elles confèrent de la résistance à la bactérie (les bactéries sont capables continuer à croître pour des concentrations de chloremphénicol plus élevé). Deuxièmement, nous pourrions utiliser des gènes "toxiques" pour la bactérie qui augmenteraient le temps de division des bactéries. De manière similaire, nous voudrions étudier extensivement la perte de plasmide car nous pensons que ce phénomène donne accès à la relation symbiotique entre les bactéries et les plasmides.

Au cours de ces "évolutions", on peut penser que les variants phénotypiques sont la source du changement de caractéristique de la population et que la fréquence de ces variants contrôle la dynamique du changement. Nous voulons vérifier cela.

## Cinquième partie

### Annexes et bibliographie

## Annexe A

# Les protéines fluorescentes utilisées

Nous avons utilisé plusieurs protéines fluorescentes au cours de cette thèse. Les trois principales furent la EGFP, la mOrange et la Venus.

**EGFP** Il s'agit d'une protéine monomérique extrêmement utilisée en biologie extraite de la méduse *Aequorea victoria*. Etant la première à avoir été découverte, elle est la mieux caractérisée des protéines fluorescentes. En effet, elle est utilisée pour des marquages de tous genres, pour des fusions, pour la reconnaissance de cellules, etc. L'ensemble de ses possibilités s'étend à presque tous les domaines de la biologie. Elle absorbe dans le bleu pour émettre dans le vert et mature vite. Le fait qu'elle soit monomérique joue beaucoup dans sa maniabilité. Celle que nous utilisons était commercialisée par Clontech. [\[71\]](#)

**mOrange** Cette protéine fluorescente est un dérivé de la DsRed et donc issu d'un corail. Son premier avantage est sa fluorescence décalée vers le rouge qui lui permet d'être distinguée de la fluorescence verte [\[51\]](#). Contrairement à ses ancêtres, la mOrange est monomérique et très fluorescentes. Nous l'avons préféré notamment à la mRFP pour son excitabilité à 488nm. D'autres protéines auraient pu être intéressants.

**Venus** La Venus est un dérivé de la YFP qui est elle-même un dérivé de la GFP. Son temps de maturation extrêmement court en fait un partenaire de choix dans toutes les applications nécessitant une visualisation en temps réel de la production de fluorescence. [\[52\]](#)

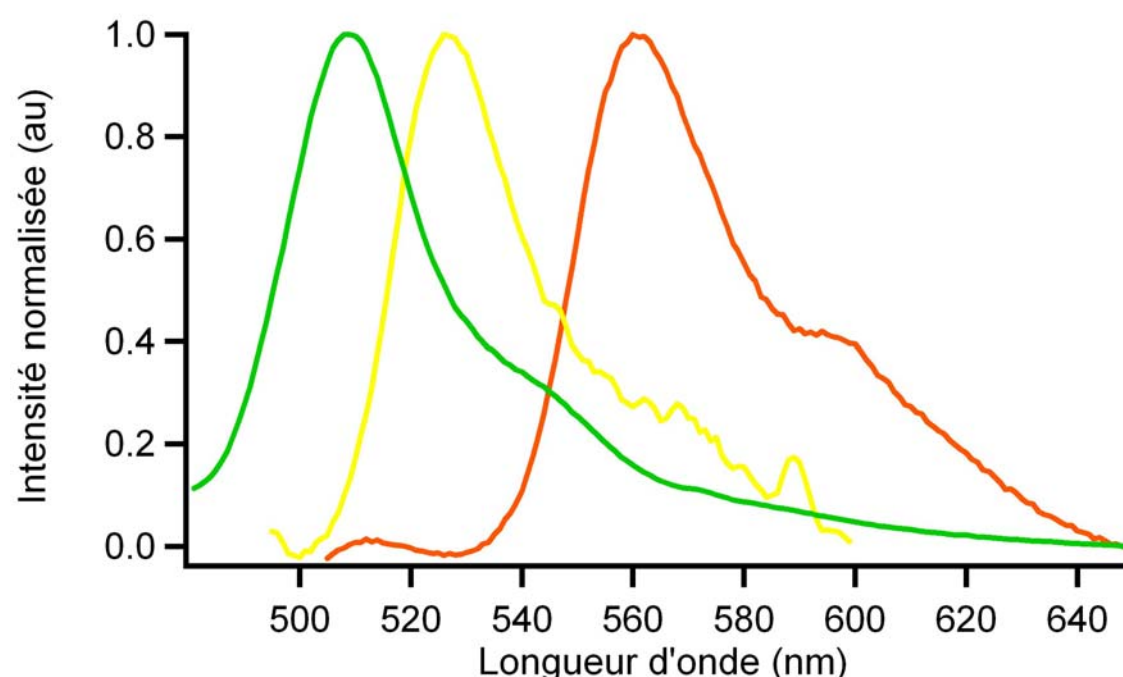


FIG. A.1: Spectres d'émission de la eGFP(vert), de la Venus(jaune) et de la mOrange (orange). Ces spectres ont été mesurés directement sur des populations liquides de bactéries. Le spectre de la venus est assez bruité. Le pic vers 590 nm est en fait un artefact lié à l'instrument.

**DsRed.T3** Cette protéine est tétramérique. Elle est surtout utilisée dans des applications de localisation. A ne pas utiliser, si on désire faire une étude quantitative à bas niveau d'expression. De plus, elle est toxique pour les bactéries. Même si elle présente[72]

**DsRed monomérique** Cette protéine présente une très petite section efficace. Ce qui la rend pratiquement inutilisable dans nos applications. Clonetech

**dTomato et tdTomato** La dTomato a de bonnes caractéristiques de fluorescence (comparables à la eGFP). Elle reste cependant un second choix pour ce qui est des protéines fluorescentes notamment à cause du fait qu'elle est dimérique. La tdTomato est en formé de deux dTomato reliés par une séquence d'acides aminés. Elle est difficile à utiliser pour deux raisons. La séquence de son gène est formée de deux gènes de dTomato. Ce qui fait qu'elle présente déjà de très fortes homologues au sein de son propre gène. Il n'est pas forcément facile compte tenu de la stratégie que nous avons adopté de faire des PCR à

bras pendants ou encore d'utiliser le kit de recombinaison. De plus, on observe la formation d'agrégats fluorescents au sein des bactéries, ce qui montre que cette protéine est non seulement toxique, mais qu'elle change le métabolisme de la bactérie. Nous voulons ici éviter cette situation. [\[71\]](#)



Protéine	$\lambda_{Amax}$ (nm)	$\lambda_{Emax}$ (nm)	$A_{488}/A_{max}$	type(n)	$EC(\text{mM}^{-1}) * QY / n = B_r$	$\tau_m$	$\tau_B(\text{s})$
EGFP	488	507	1	monomer(1)	$56 * 0.6 / 1 = 34$	< 1h	174
mOrange	548	562	$\sim 0.25$	monomer(1)	$71 * 0.69 = 49$	2.5h	6.4 ou 9
Venus	515	528	$\sim 0.25$	monomer(1)	$92 * 0.57 = 52$	< 1min	15
DsRed.T3	560	587	$\sim 0.5$	tetramer(4)	$49.5 * 0.59 / 4 = 7.3$	1.3h	??
dTomato	554	581	$\sim 0.25$	dimer(2)	$69 * 0.69 / 2 = 23.75$	1h	64
tdTomato	554	581	$\sim 0.25$	tandem dimer(1)	$138 * 0.69 / 1 = 95$	1h	70

TAB. A.1: Ce tableau résume quelques caractéristiques intéressantes concernant les protéines que nous avons utilisé. Les données compilées ont été soit mesuré, ou extrait de la bibliographie [71] [51] [52] [72] .

$\lambda_{Amax}$  est la longueur d'onde d'excitation maximum.

$\lambda_{Emax}$  est la longueur d'onde d'émission maximum.

$A_{488}/A_{max}$  est l'efficacité de l'excitation à 488 par rapport à l'excitation maximum.

$n$  est le nombre de protéines nécessaires pour obtenir une protéine fonctionnelle.

$EC$  est le coefficient d'extinction. Cette grandeur qualifie l'efficacité qu'a la protéine à absorber des photons.

$QY$  est le quantum yield qui lui traduit l'efficacité de l'émission.

$B_r$  est la brillance par protéine.

$\tau_m$  est le temps nécessaire à la maturation de la moitié des protéines produites.

$\tau_B$  est le temps de bleach.

# Annexe B

## Protocoles

### B.1 Biologie moléculaire et Microbiologie

La plupart des protocoles rappelés dans cette section sont directement issu d'un manuel d'expérimentation [73]. Il s'agit ici de rappels et d'aide mémoire plus qu'autre chose.

#### B.1.1 Stocks et Solutions

**LB Broth** 10g/L Tryptone  
5g/L Yeast extract  
10g/L NaCl

**LB LowSalt** 10g/L Tryptone  
5g/L Yeast extract  
5g/L NaCl

**LB Agar** 10g/L Tryptone  
5g/L Yeast extract  
10g/L NaCl  
15g/L Bacteriological Agar

**Phosphate Buffer Saline (PBS)** 2,7mM (0,2g/L) KCl  
137mM (8g/L) NaCl  
2mM (0,24g/L) KH<sub>2</sub>PO<sub>4</sub>  
10mM (1,44g/L) Na<sub>2</sub>HPO<sub>4</sub>

Produit	Concentration du Stock	Remarque
IPTG	1M	Différent de la concentration de <a href="#">[73]</a>
Ampicilline	50mg/mL	
Streptomycine	10mg/mL	
kanamycine	10mg/mL	
Chloremphénicol	34mg/mL	dans éthanol
tétracycline	5mg/mL	se dégrade à la lumière, dans éthanol

### B.1.2 Protocoles de biologie moléculaire

**Restriction** Mettre en solution le buffer de l'enzyme, l'enzyme et l'ADN à digérer. Placer à 37°C. La plupart des enzymes que nous avons utilisés reconnaissent des sites palindromiques et n'ont pas de Star Activity. Ils proviennent en grande partie de chez Fermentas.

**Ligation/Circularisation** Mettre en solution le buffer de la ligase, la ligase, et les morceaux d'ADN à liguer ou à circulariser. Placer à 4°C overnight. Faire plusieurs rapports entre les différents fragments. Déterminer l'optimum empiriquement.

**Polymérase Chain Reaction (PCR)** Mettre en solution ADN à amplifier, oligonucléotides, TAC polymérase, dNTPs et buffer de la TAC. Placer dans la machine à PCR, configurer les températures du cycle et les temps d'attentes. Le design des oligos semble crucial. Utiliser [primer3](#) pour les d. Pour des PCR supérieure à 2,5kb utiliser d'autres polymérases peut aider, notamment lorsque les cycles se font avec des incréments de temps. Toujours purifier à la fin au moins sur colonne

**purification colonne/gel** Nous utilisons le kit [Qiaquick Spin Column](#).

**précipitation éthanol** Rajouter 2 volume d'éthanol. Mélanger, mais ne surtout pas **vortexer**. Centrifuger à froid à 13000 RPM pendant au moins 30 min. On peut aussi rajouter du Sodium Acetate (afin de visualiser le culot d'ADN). Enlever le surnageant. Rincer le culot une deuxième fois à l'éthanol 70% sans défaire le culot. Recentrifuger et retirer le surnageant de nouveau. Laisser sécher sous une hote.

**Electroporation** Porter la souche à transformer jusqu'à  $OD_{600} \sim 0,4 - 0,5$ . Les laver deux fois à l'eau froide (Centrifuger à 9500 RPM pendant 10 min, retirer le surnageant, rajouter l'eau). Transférer 1ml de bactérie dans un eppendorf, centrifuger à 9500 RPM à 4°C, retirer le surnageant, rajouter de l'eau pour finir dans un volume total de 20 $\mu$ l. Rajouter 1 à 5  $\mu$ l d'ADN. Transférer dans une cuve d'électroporation (1mm de gap ou d'entrefer). Appliquer une tension de 1,35kV avec 600 ohms de résistance. Transférer 1ml de LB chaud à 37°C SANS antibiotiques. conserver à 37°C pendant 1h. Etaler 200 – 400 $\mu$ l sur une plaque de LB-Agar contenant les antibiotiques appropriés. La placer à la température adéquate. Prélever les colonies qui auront poussées et les tester.

**Transformation chimique** Le protocole utilisé est extrait de la 1ère édition du Molecular Cloning [73]. Porter la souche à transformer jusqu'à  $OD_{600} \sim 0,4$ . Prélever 50ml et les placer sur glace 10minutes. Centrifuger 10' à 2700g. Vider le surnageant et attendre 1' renversé. Rajouter 30ml  $CaCl_2$  à 0,1M. Centrifuger 10' à 2700g. Vider le surnageant et attendre 1' renversé. Rajouter 2ml  $CaCl_2$  à 0,1M. Prélever 200 $\mu$ l et y ajouter le produit de ligation (tester plusieurs quantités entre 1 et 10 $\mu$ l). Placer au bain-marie à 42°C pendant 90'' puis placer le produit 1 – 2' sur glace. Ajouter 800 $\mu$ l de LB sans antibiotique et placer à 37°C pendant 45'. Etaler sur la boîte sélective (avec les antibiotiques nécessaires).

**Extraction** Utiliser les kits d'extraction de plasmide [mini\(Fermentas\)](#), [midi\(Sigma\)](#) ou [maxi\(sigma\)](#)

**gel agarose** 1% permet de séparer des bandes entre 200 et 2000 bases. C'est la gamme d'ADN qu'on utilise. Pour 400ml d'eau, rajouter 2,8g d'agarose, 4ml de TAE 50x, 20 $\mu$ l BET. Fondre le gel, le mouler avec les puits. démouler, placer le gel dans une cuve à gel, utiliser le même buffer (TAE 0,5x). Appliquer une tension afin de faire migrer l'ADN. 500V pendant 10min déplacement de 3 cm pour un morceau de 1kB (ne peut se faire qu'à froid).

### B.1.3 Protocoles détaillés des constructions effectuées

#### SYSBRG

1. Extraction du plasmide clontech pEGFP (Ce plasmide n'est plus commercialisé).

2. Restriction EcoRI/HindIII du plasmide pEGFP.
3. Purification sur gel de la bande correspondant au gène EGFP (environ 750 bp).
4. Pcr avec amorces ptacI-gfp/morV2 et eGFP/xxTomato-TR2\* sur une matrice plasmide pEGFP-clontech.
5. Purification sur gel et kit [Qiaquick \(Qiagen\)](#) du produit de pcr V1.
6. Pcr avec amorces PtacI\* et TR2\* sur la matrice V1.
7. Purification sur gel et kit [Qiaquick \(Qiagen\)](#) du produit de pcr V2. V2 est un fragment d'ADN composé des sites de restriction EcoRI et XhoI, d'un promoteur Tac, du gène EGFP, d'une séquence de terminaison de transcription et du site de restriction HindIII.
8. Pcr avec les bras d'homologie, amorces homotac2 et WG2004-04 sur la matrice V2.
9. Purification sur gel kit [Qiaquick \(Qiagen\)](#) du produit de pcr V3.
10. Electroporation de l'ADN V3 dans des bactéries TOP10-lacIq portant le plasmide de recombinaison genebridge.
11. Vérification de l'insertion par séquençage (amorces pyrFup et pyrFlow) du produit de pcr (amorces pyrFup et pyrFlow) directement sur les bactéries transformées.

### pUC-BRO

1. Extraction du plasmide pRSET-BmOrange (Tsien).
2. Restriction BamHI/EcoRI du plasmide extrait.
3. Purification sur gel de la bande correspondant au gène de la mOrange (environ 750 pb).
4. Pcr avec amorces ptacI-gfp/morV2 et eGFP/xxTomato-TR2\* sur la matrice extraite précédemment.
5. Purification sur gel et kit [Qiaquick \(Qiagen\)](#). du produit de pcr O1.
6. Pcr avec amorces PtacI\* et TR2\* sur la matrice O1.
7. Purification sur gel et kit [Qiaquick \(Qiagen\)](#). du produit de pcr O2. O2 est un fragment d'ADN composé des sites de restriction EcoRI, XhoI d'un promoteur Tac, du gène mOrange, d'une séquence de terminaison de transcription et du site de restriction HindIII.

8. Restriction EcoRI/HindIII de l'ADN O2.
9. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction O2.
10. Restriction EcoRI/HindIII du plasmide pUC19.
11. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction du plasmide pUC19.
12. Ligation des fragments cohésifs de l'ADN O2 et du pUC19.
13. Transformation de bactéries DH5 $\alpha$  compétentes au calcium avec le produit de ligation.
14. Vérification de l'insertion par séquençage (amorces universelles -21 et -29) du plasmide extrait des clones.

## **SYSBRO**

1. Extraction du plasmide pUC-BRO.
2. Restriction XhoI/HindIII du plasmide extrait.
3. Purification sur gel de la bande correspondant au fragment d'ADN O2 (environ 900 pb).
4. Pcr avec les bras d'homologie, amorces homotac2 et WG2004-04 sur la matrice O2.
5. Purification sur gel kit [Qiaquick \(Qiagen\)](#). du produit de pcr O3.
6. Electroporation de l'ADN O3 dans des bactéries TOP10-lacIq portant le plasmide de recombinaison genebridge.
7. Vérification de l'insertion par séquençage (amorces pyrFup et pyrFlow) du produit de pcr (amorces pyrFup et pyrFlow) directement sur les bactéries transformées.

## **pZC-BRO**

1. Extraction du plasmide pUC-BRO.
2. Restriction XhoI/HindIII du plasmide extrait.
3. Purification sur gel de la bande correspondant au fragment d'ADN O2 (environ 900 pb).
4. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction O2.

5. Restriction XhoI/HindIII du plasmide pZC320 (voir chapitre 4).
6. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction du plasmide pZC320.
7. Ligation des fragments cohésifs de l'ADN O2 et du pZC320.
8. Transformation de bactéries DH5 $\alpha$  compétentes au calcium avec le produit de ligation.
9. Vérification de l'insertion par séquençage (amorces morangefin et FcHind) du plasmide extrait des clones.

### **pBR-BRO**

1. Extraction du plasmide pUC-BRO.
2. Restriction EcoRI/HindIII du plasmide extrait.
3. Purification sur gel de la bande correspondant au fragment d'ADN O2 (environ 900 pb).
4. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction O2.
5. Restriction EcoRI/HindIII du plasmide pBR322 (voir chapitre 4).
6. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction du plasmide pBR322.
7. Ligation des fragments cohésifs de l'ADN O2 et du pBR322.
8. Transformation de bactéries DH5 $\alpha$  compétentes au calcium avec le produit de ligation.
9. Vérification de l'insertion par séquençage (amorce morangefin) du plasmide extrait des clones.

### **pOU-BRO**

1. Extraction du plasmide pUC-BRO.
2. Restriction EcoRI/HindIII du plasmide extrait.
3. Purification sur gel de la bande correspondant au fragment d'ADN O2 (environ 900 pb).
4. Pcr avec les amorces Ptacl\* et TR2\_pagI de la matrice O2.

5. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de pcr O2'.
6. Restriction EcoRI/NcoI (ou pagI, à vérifier) de l'ADN O2'.
7. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction O2'.
8. Restriction EcoRI/NcoI du plasmide pOU82 (voir chapitre 4).
9. Purification sur gel avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction du plasmide pOU82.
10. Ligation des fragments cohésifs de l'ADN O2' et du pOU82.
11. Transformation de bactéries DH5 $\alpha$  compétentes au calcium avec le produit de ligation.
12. Vérification de l'insertion par séquençage (amorce morangefin) du plasmide extrait des clones.

### **pOAR-BRO**

1. Extraction du plasmide pUC-BRO.
2. Restriction EcoRI/HindIII du plasmide extrait.
3. Purification sur gel de la bande correspondant au fragment d'ADN O2 (environ 900 pb).
4. Pcr avec les amorces PtacI\* et TR2\_pagI de la matrice O2.
5. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de pcr O2'.
6. Restriction EcoRI/NcoI (ou pagI, à vérifier) de l'ADN O2'.
7. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction O2'.
8. Restriction EcoRI/NcoI du plasmide pOAR64 (voir chapitre 4).
9. Purification sur gel avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction du plasmide pOAR64.
10. Ligation des fragments cohésifs de l'ADN O2' et du pOAR64.
11. Transformation de bactéries DH5 $\alpha$  compétentes au calcium avec le produit de ligation.
12. Vérification de l'insertion par séquençage (amorce morangefin) du plasmide extrait des clones.



**pZC-O+V**

1. Construction des fragments V2 et O2 tels que décrit précédemment.
2. Restriction XhoI des fragments V2 et O2.
3. Purification avec le kit [Qiaquick \(Qiagen\)](#). des produits de restriction.
4. Ligation des fragments cohésifs V2 et O2.
5. Purification sur gel avec le kit [Qiaquick \(Qiagen\)](#). de la bande d'ADN correspondant au dimère V2+O2 (environ 1800 pb). V2+O2 est un fragment d'ADN composé du site de restriction HindIII, d'une séquence de terminaison de transcription, du gène EGFP, de 2 promoteurs Tac "tête bêche", d'un gène mOrange et d'une séquence de terminaison.
6. Restriction HindIII du plasmide pZC320.
7. Purification avec le kit [Qiaquick \(Qiagen\)](#). du produit de restriction.
8. Ligation des fragments cohésifs pZC320 et V2+O2.
9. Transformation de bactéries DH5 $\alpha$  compétentes au calcium avec le produit de ligation.
10. Vérification de l'insertion par séquençage (amorce bamsense) du plasmide extrait des clones.

Listes des amorces de pcr									
ptacI-gfp/morV2	atgtgtggaa	tttgagcgg	ataacaattt	cacacaggag	gtaagatcta	tggtgagcaa	gggcgaggag		
eGFP/xxTomato-TR2*	aaaggcctgc	gattaccagc	aggcctgttc	ttacttgtac	agctcgcca	tgcc			
PtacI*	acctgtgaat	tctcgagcta	gcgttgacaa	ttaatcatcg	gctcgtataa	tggtgtggaat	tgtagcgcca		
TR2*	accagtaagc	ttaagcatgc	gtctaccaaa	taaaaaggcc	tgcgattacc	agc			
homotac2	actaattatt	aatgttccat	tgtgtccgg	caacgacggc	gcagaaaaag	cctgccaggg	ttctcgagct		
	agcgttgaca	atta							
WG2004-04	aacaggtaaa	cggcgagcat	tctaccgtg	gaaataaaaa	attaacagtt	gcgatttctc	gtccatggta		
	ccaataaaaa	aggcctgcga							
PyrFup	atgttccatt	gtgtccggc	aacg						
PyrFlow	cggcgagcat	tctaccgtg	gaaat						
morangefin	ttacttgtac	agctcgcca	tgcc						
FcHind	ataataaagc	ttacgatcat	cgtgcgggag	ggc					
TR2_pagI	accagtaagc	ttcatgagct	cttaagcatg	cgtctacca	ataaaaaaggc	ctgcgattac	cagcagcct		
	gttc								
Bamsense	gccgatcca	cgtgatcg							

TAB. B.1: Les oligos nucléotides présentés ici ont été mentionnés à divers moment du manuscrit

## B.2 Cellules microfluidiques

### B.2.1 Fabrication des moules

1. Dégraisser les lames de verre au toluène pendant 30 min au soniqueur.
2. Enlever toutes traces organiques du verre au "piranha" (50% peroxyde 30% et 50% acide sulfurique)
3. Rincer les lames, les sécher, les porter à 200°C pendant 10 min, puis les ramener à 100°C.
4. Etaler la résine SU8-2002 en spincoatant à 3000RPM pendant 30s.
5. Précuire à 95°C pendant 1 min
6. Placer le masque et insoler à 4 minutes aux UV
7. Cuire à 95°C pendant 2 min
8. Révéler le motif en rinçant délicatement au SU8-remover
9. Rincer à l'isopropanol (attention à ne pas décoller la résine polymérisée)
10. Coller un anneau métallique

### B.2.2 La cellule

1. Mélanger le RTVa et RTVb dans les proportions 10 :1 (22g de A et 2.2g de B pour 12 moules)
2. Agiter fermement afin d'homogénéiser les produits. Centrifuger afin de retirer la majeure partie des bulles
3. Dégaser le PDMS en le mettant sous vide. Centrifuger afin de retirer les bulles
4. Verser le PDMS dans les moules et faire cuire 2h à 90°C (des temps plus longs ne sont absolument pas gênants) La réticulation à température est très lente (plusieurs jours)
5. Laver les lamelles de quartz ou de verre selon le même protocole que les lames de verre
6. Démouler le PDMS et le retourner

7. Découper le PDMS de manière à obtenir des réservoirs au bout de chaque canal. Retirer le PDMS superflu pour les lamelles de quartz
8. Placer le PDMS et une lamelle au plasma cleaner (5 secondes)
9. Déposer délicatement le PDMS, il devrait se coller au verre immédiatement. (attention aux morceaux de PDMS coincés)
10. Remplir les réservoirs de PBS-Chloremphénicol (ne pas tarder, le faire tant que la surface interne des canaux est encore hydrophile).

### B.2.3 Préparation des bactéries

Les bactéries croissent toujours dans des milieux supplémenté avec les antibiotiques nécessaires.

1. Etaler la souche d'intérêt sur une boîte de LB-agar. Placer dans un incubateur à la température désirée toute la nuit (over night).
2. Prélever une colonie unique pour inoculer 2ml de LB dans un tube falcon de 50ml. Placer dans un incubateur à la température désirée en agitant à 250RPM pour l'aération et ce toute la nuit (over night).
3. Inoculer un erlenmeyer de 50ml de LB. Placer dans un incubateur à la température désirée en agitant à 250RPM.
4. Prélever 10ml lorsque la solution est à la bonne OD<sub>600</sub>. Séparer en deux tubes falcon de 5ml bouchés avec de la mousse. Pour mesurer l'OD<sub>600</sub>, faire deux prélèvement de 1ml et faire la mesure plusieurs fois à plusieurs dilutions (1 , 2 et 4 fois).
5. Induire un des tubes à 1mM d'IPTG en concentration finale. Remettre les tubes à la même température avec la même agitation.
6. Bloquer la production de protéines dans les deux tubes en rajoutant du chloremphenicol à 170 $\mu$ g/ml au bout de 20' d'induction.
7. Laver les bactéries au PBS supplémenté de chloremphenicol à 170 $\mu$ g/ml (PBS-Chlo) après 1h de maturation dans le même tube à la température désirée en agitant à 250RPM. Pour laver les bactéries, il faut centrifuger à 6000RPM à 4°C pendant 10', retirer le surnageant en laissant le tube renversé pendant 1' et finalement rajouter le même volume de PBS-Chlo à 170 $\mu$ g/ml. Ne pas vortexer.
8. Placer à 37°C pour maturation toute la nuit (over night) dans des tubes bouchés en agitant à 250RPM. Ensuite, retirer les tubes de l'incubateur et les placer à 4°C dans de l'aluminium pour éviter le photo-blanchiement des protéines fluorescentes.
9. Laver les bactéries au PBS-Chlo mais ressuspendre dans un volume de 0,5ml.
10. Mélanger les bactéries induites et les non-induites (typiquement dans un rapport 5 pour 1) puis filtrer la solution finale à 5 $\mu$ m pour éliminer les agrégats de bactéries.

## Annexe C

# Analyse de données

L'acquisition des données donne la valeur de la fluorescence détectée par chacune des photodiodes avalanches en fonction du temps. Comme le flux dans les canaux n'est pas très bien contrôlé, l'échantillonnage se fait entre 1 et 8kHz dépendant de la vitesse de défilement des bactéries dans le champ d'observation. Les enregistrements typiques présentent une ligne de base bruitée modifiée par l'arrivée de bactéries dans le champ d'acquisition des APDs qui produit un créneau dans le signal enregistré.

Ces séries temporelles sont analysées à l'aide de procédures développées sous IgorPro5 et Matlab :

1. Sur chacune des traces, extraction du moment et de l'intensité des événements. Ajustement des paramètres nécessaires.
2. Regroupement des événements survenus au même moment.
3. Affichage d'un diagramme de corrélation (Fluorescence verte et fluorescence rouge) et classification manuelle (manual clustering) afin d'isoler des sous populations si nécessaire.
4. Construction des diverses représentations nécessaires (histogramme, calcul des valeurs statistiques intéressantes).

**Détection des événements.** Nous recherchons les fortes variations soutenues. Pour cela, à partir de la même trace temporelle de la fluorescence, nous construisons deux traces *Baseline* et *Peak* et qui sont des moyennes glissantes dont la fenêtre de lissage peut être différente. L'une sera notre référence pour le niveau de base, l'autre pour la

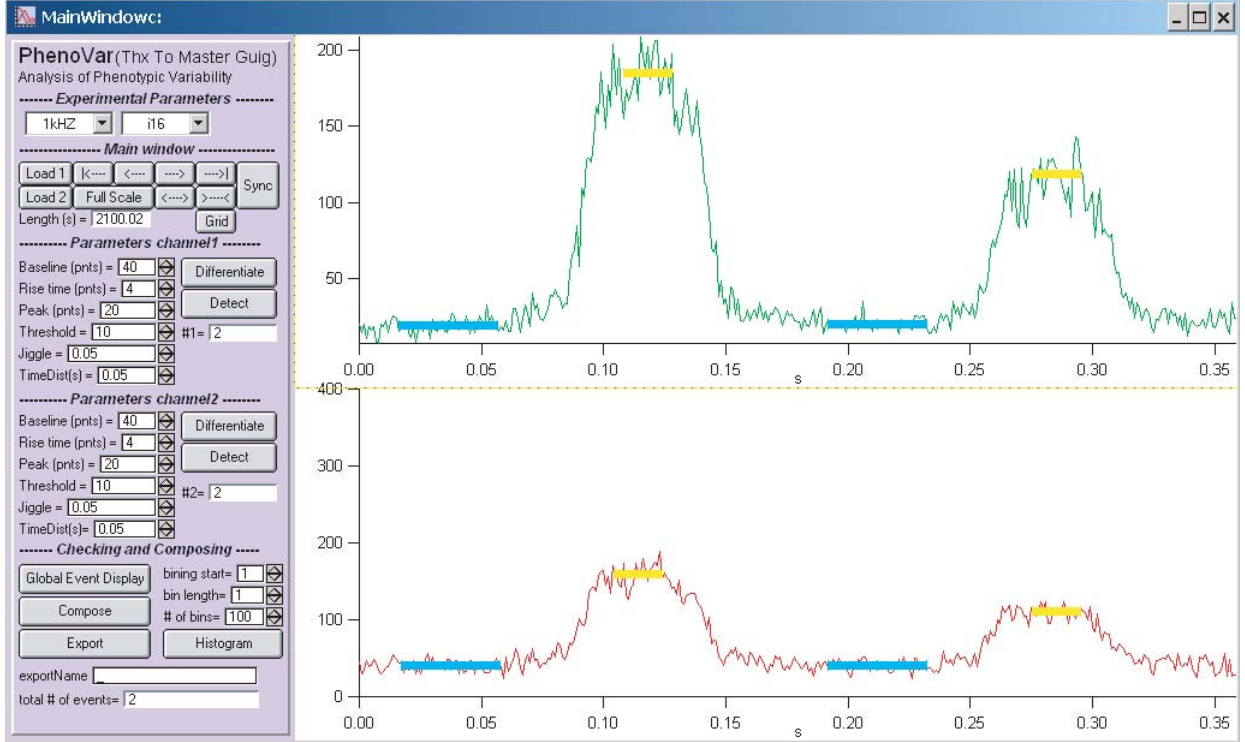


FIG. C.1: L'interface d'analyse développée sous IgorPro 5 permet une navigation rapide et conviviale des données brutes. On peut ainsi faire varier plusieurs paramètres de détection

hauteur du créneau. A partir de ces deux vecteurs, on peut en construire un troisième qui nous donnera une sorte de dérivée. Pour cela, on soustrait à une valeur de *Peak*, une valeur légèrement antérieure dans *Baseline*. L'écartement d'indice entre ces deux valeurs peut être ajustée. On obtient alors un genre de dérivée lissée, *Diff*.

$$\begin{aligned}
 Baseline(i) &= \sum_{j=-N_B/2}^{N_B/2} Trace(i+j) \\
 Peak(i) &= \sum_{j=-N_P/2}^{N_P/2} Trace(i+j) \\
 Diff(i) &= Peak(i+N_R) - Baseline(i)
 \end{aligned}$$

$i$  et  $j$  sont des indices temporels.  $N_B$  et  $N_P$  sont les tailles des fenêtre de lissage pour *Baseline* et *Peak*.  $N_P$  est choisi de manière à être un peu plus court que le créneau.  $N_B$  est choisi le plus grand possible sans que sa taille n'excède l'espace minimum entre deux bactéries.  $N_R$  est l'espacement utilisé pour construire *Diff*.  $N_R$  sera de l'ordre du temps de montée du créneau.

Il suffit alors de détecter les moments auxquels *Diff* passe au dessus d'un seuil ajustable pour localiser les évènements. En soustrayant une valeur antérieure de *Baseline* au maximum de *Peak* autour de la localisation de l'évènement, on obtient l'amplitude de l'incrément lié à la bactérie, *RealAmp*.

**Incertitude sur l'intensité de fluorescence des bactéries** Lorsqu'on se donne une trace, la variance sur la hauteur moyenne du créneau  $\delta P^2$  est causée par le bruit de photon. On peut alors écrire  $\delta P^2 = P/N_P$  où  $P$  est la hauteur moyenne du créneau au sein de la fenêtre de lissage et  $N_P$  est la taille de la fenêtre de lissage. La variance sur *RealAmp* peut s'écrire :

$$\delta R^2 = \frac{\delta P^2}{N_P} + \frac{\delta B^2}{N_B}$$

Dans le cas de bactéries que l'on a préparé sans induction (dont la fluorescence est extrêmement basse), on a une incertitude moyenne de détermination sur le canal vert  $\sqrt{\langle \delta R_g^2 \rangle} = 0.3373u.a.$  alors que la moyenne des évènements  $\langle R_g \rangle = 5.9766u.a.$ ; soit  $\sqrt{\langle \delta R_g^2 \rangle} / \langle R_g \rangle \sim 6\%$ . Par contre, pour des bactéries exprimant la eGFP à partir d'un gène situé sur un chromosome, on obtient  $\sqrt{\langle \delta R_g^2 \rangle} / \langle R_g \rangle \sim 3\%$ .

Lorsqu'on compare  $\langle \delta R_g^2 \rangle$  à l'écart-type de la fluorescence dans une population de bactéries non-induites, on trouve  $\sqrt{\langle \delta R_g^2 \rangle} / \sqrt{\langle R_g^2 \rangle - \langle R_g \rangle^2} \sim 13\%$ . Pour une population de bactéries exprimant la eGFP à partir d'un gène situé sur le chromosome, nous avons obtenu  $\sqrt{\langle \delta R_g^2 \rangle} / \sqrt{\langle R_g^2 \rangle - \langle R_g \rangle^2} \sim 7\%$ .

**Regroupement des évènements.** On compare les temps d'occurrence des évènements sur le premier et le deuxième détecteur. Pour tous les évènements n'ayant pas de correspondant sur l'autre détecteur, on lui en rajoute un de valeur nulle.

**Classification** Nous avons opté pour une classification manuelle<sup>1</sup> car les méthodes automatiques deviennent réellement fiable lorsque le nombre de variables est supérieur à deux. Dans notre cas, nous ne mesurons que la fluorescence verte et la fluorescence rouge.

La séparation des sous-population permet d'extraire les paramètres statistiques pertinents pour chacune des sous populations. Ne pas les séparer rendrait l'analyse dépendant

---

<sup>1</sup>Nous avons directement utilisé le script suivant :

<http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=13857>



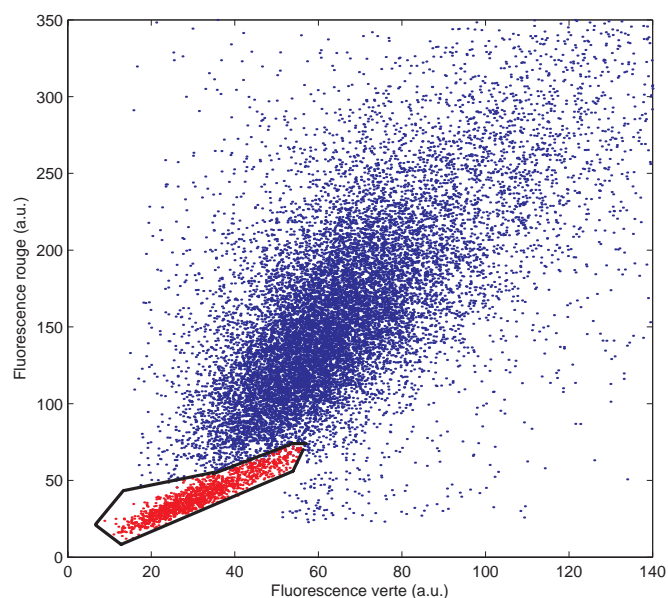


FIG. C.2: A l'aide d'un script Matlab, on sélectionne les points d'intérêt au sein d'un diagramme de corrélation. Ici, une sous-population en rouge a été extraite. Il s'agit d'une population non-induite.

des proportions des différentes populations et ne donneraient pas d'estimations correctes.

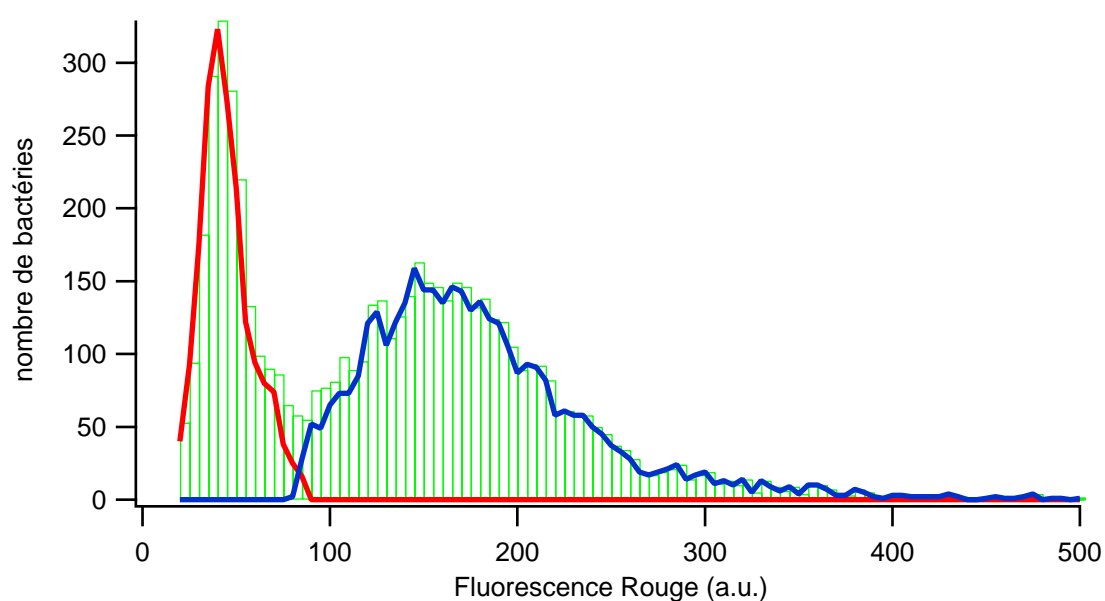


FIG. C.3: A partir des sous-populations sélectionnées à l'étape précédente, on est capable d'extraire l'histogramme des fluorescence. En vert, l'histogramme de fluorescence sur un canal de toute la population mesurée. En rouge et bleu, l'histogramme des population rouge et bleu de la figure [C.2](#)

## Annexe D

# Incertitude statistique et incertitude de mesure

**Incertitude statistique** Soit une variable aléatoire  $x$  dont les réalisations se font à partir de la distribution  $L$ , définie par ses moments  $\langle x^n \rangle$ . Le meilleur estimateur de  $\langle x \rangle$  est la variable aléatoire  $m = \frac{1}{N} \sum_{i=1}^N x_i$ . Tous les  $x_i$  tirent leur réalisations de la distribution  $L$ . Il est possible de calculer  $\sigma_m^2$ , la variance de  $m$ . On peut alors écrire :

$$\langle x \rangle = m \pm \sqrt{\sigma_m^2} \quad \text{avec} \quad \sigma_m^2 = \frac{\langle x^2 \rangle - \langle x \rangle^2}{N}$$

Nous prendrons comme estimateur de la variance  $v = \frac{1}{N-1} \left( \sum_i x_i^2 - \frac{1}{N} (\sum_i x_i)^2 \right)$ . De la même manière, nous pouvons calculer  $\sigma_v^2$ , la variance sur  $v$ .

$$\langle x^2 \rangle - \langle x \rangle^2 = v \pm \sqrt{\sigma_v^2} \quad \text{avec}$$

$$\sigma_v^2 = \frac{1}{N} \left( \langle x^4 \rangle - 4\langle x^3 \rangle \langle x \rangle - \frac{n-2}{n-1} \langle x^2 \rangle^2 + \frac{8n-11}{n-1} \langle x^2 \rangle \langle x \rangle^2 - \frac{4n-6}{n-1} \langle x \rangle^4 \right)$$

**Considérer l'incertitude de mesure** Lors de notre mesure, nous n'avons pas accès à  $x$  mais à  $X = x + \delta x$ . Considérons la variable aléatoire  $X$  de distribution  $L'$ . On utilise alors l'estimateur  $m' = \frac{1}{N} \sum_i (X_i) = \frac{1}{N} \sum_i (x_i + \delta x_i)$ . Supposant que  $\sum_i \delta x_i = 0$ , on obtient  $m' = m$ . Ainsi,

$$\langle x \rangle = m' \pm \sqrt{\sigma_{m'}^2} \quad \text{avec} \quad \sigma_{m'}^2 = \frac{\langle X^2 \rangle - \langle X \rangle^2}{N} \quad (\text{D.1})$$

L'incertitude sur  $\langle X \rangle$  prend donc en compte l'incertitude de mesure que l'on peut avoir sur  $x$ .

Lorsqu'on considère  $v' = \frac{1}{N-1} \left( \sum_i X_i^2 - \frac{1}{N} (\sum_i X_i)^2 \right)$ , nous avons supposé les moyennes faisant intervenir puissances impaires de  $\delta x$  nulles (e.g.  $\langle x \delta x \rangle = 0$ ). On obtient alors :

$$v' = v + \langle \delta x^2 \rangle$$

Donc  $v'$  surestime  $\langle x^2 \rangle - \langle x \rangle^2$  de  $\langle \delta x^2 \rangle$ . Dans nos systèmes, nous avons constaté que la variance sur la moyenne des  $\delta x^2$  est 1 à 2 ordre de grandeur plus petit que  $\sigma_{v'}^2$ , on peut alors écrire :

$$\langle x^2 \rangle - \langle x \rangle^2 \simeq v' - \langle \delta x^2 \rangle \pm \sqrt{\sigma_{v'}^2} \quad \text{avec} \quad (D.2)$$

$$\sigma_{v'}^2 = \frac{1}{N} \left( \langle X^4 \rangle - 4 \langle X^3 \rangle \langle X \rangle - \frac{n-2}{n-1} \langle X^2 \rangle^2 + \frac{8n-11}{n-1} \langle X^2 \rangle \langle X \rangle^2 - \frac{4n-6}{n-1} \langle X \rangle^4 \right)$$

Les équations D.1 et D.2 font intervenir les véritables moments de la distribution L dans le calcul de l'incertitude. Ainsi, dans notre cas, lors du calcul de  $\sigma_{m'}^2$  et  $\sigma_{v'}^2$ , nous utiliserons  $\frac{1}{N} \sum_i X^n$  au lieu de  $\langle X^n \rangle$ .

## Annexe E

### Article soumis à PRE

Cet article a été soumis après la rédaction et la soutenance de la thèse. Certaines données n'apparaissent que dans cet article.

# Plasmid copy number noise in monoclonal populations of bacteria

Jérôme Wong Ng,<sup>\*</sup> Didier Chatenay, and Jérôme Robert<sup>†</sup>

*Laboratoire Jean Perrin, FRE 3231 CNRS-UPMC,*

*24 rue Lhomond, 75005 Paris, France.*

Michael Guy Poirier

*Department of Physics, Ohio State University*

*191 W Woodruff Ave, Columbus, OH 43210 , USA*

(Dated: October 27, 2009)

## Abstract

Plasmids are extra chromosomal DNA that can confer to their hosts supplementary characteristics such as antibiotic resistance. Plasmids code for their copy number through their own replication frequency. Even though the biochemical networks underlying the plasmid copy number (PCN) regulation processes have been studied and modeled, no measurement of the heterogeneity in PCN within a whole population has been done.

We have developed a fluorescent-based measurement system which enables determination of the mean and noise in PCN within a monoclonal population of bacteria. Two different fluorescent protein reporters were inserted, one on the chromosome and the other on the plasmid. The fluorescence of these bacteria was measured with a microfluidic flow cytometry device. We show that our measurements are consistent with known plasmid characteristics. We find that the partitioning system lowers the PCN mean and standard deviation. Finally, bacterial populations were allowed to grow without selective pressure. In this case, we were able to determine the plasmid loss rate and growth inhibition effect.

---

<sup>\*</sup>Currently at Unité génétique In Silico, Institut Pasteur. 25-28, rue du Dr Roux, 75015 Paris, France.

<sup>†</sup>Electronic address: [jerome.robert@lps.ens.fr](mailto:jerome.robert@lps.ens.fr)

## I. INTRODUCTION

Plasmids are extra chromosomal DNA fragments that constitute as much as 25% of some organisms' genetic material [1]. In bacteria, the genes coded on plasmids often transfer between different species making them an interesting pool of shared genetic material [2]. In addition to genes expressed on the chromosome, plasmids provide many phenotypes to bacteria. Some of them are vital to the bacterium such as antibiotic resistance or toxin-antitoxin systems [3]. In order to replicate themselves, plasmids use the replication machinery of their hosts. These features allow the plasmids to exhibit interesting symbiotic behaviors with their host. Moreover, plasmids code for a biochemical network that regulates their replication frequency [4]. This regulation network, which varies from plasmid to plasmid, sets the plasmid copy number (PCN). Researchs on plasmids have focused on unraveling and understanding the networks controlling the replication frequency [4]. For example, the use of non-coding RNA in regulation processes were first discovered in plasmids [5].

Fluctuations in plasmid replication frequency and timing or allocation of different copy numbers at cell division can cause the PCN to vary from bacteria to bacteria within an isogenic population. Within this framework, some quantitative models of PCN regulation evaluate the PCN distribution [6–8].

In order to understand these differences in PCN, it is important to determine both the mean and standard deviation of the PCN. On the experimental side, DNA titration is not a very precise measurement of the mean PCN. Quantitative PCR (qPCR) may be more accurate [9] but does not give access to the standard deviation of the PCN when performed on a population. To our knowledge, no measurement of the PCN distribution is available in the literature.

In order to measure the PCN distribution in a population, many individual cell measurements of the PCN within a large bacterial population are required. One solution is *in vivo* labeling techniques. For example, insertion of a *lac* array into the plasmid and GFP-LacI fusion expression leads to fluorescent focal points [10] that can be counted. While these techniques proved suitable for extracting localization dynamics of chromosomes, they are problematic in cases involving higher numbers of objects for two reasons. First, the optical diffraction limits the number of distinguishable objects within a confined volume such as

the bacterium cytoplasm. Second and more importantly, in cases of plasmids with low copy numbers between 5 – 10, only 2 – 6 focal points could be observed [11], which indicates that either plasmids spontaneously form clusters or fluorescent probes aggregate. In both cases, quantification of the PCN by focal point measurements will be erroneous.

Another approach is to measure the expression of a gene inserted in the plasmid to extract the PCN. However, both the PCN distribution and the noise in gene expression contribute to the expression distribution[12]. In order to diminish the noise sources, our strategy is to insert a second gene in the bacterial chromosome. With the exceptions of the noise in PCN and chromosomal copy number, the expression from this chromosomal gene is submitted to the same sources of noise as the plasmid gene. Hence, the expression distribution from a plasmid can be corrected by the expression noise from the chromosome, which allows us to determine information about the PCN distribution. More precisely, we constructed an *E. Coli* strain in which the *egfp* [13] gene was inserted into the chromosome opposite to the chromosome *ori* to minimize copy number fluctuations during the cell cycle. The *mOrange* [14] gene was inserted in each of the studied plasmids. The *egfp* bearing strain was then transformed with each plasmid separately. We chose the *mOrange* protein because it is the brightest fluorescent protein within the orange-red spectrum. Moreover, it could be reasonably excited with the same wavelength as the EGFP. Both genes were under control of the same isopropyl-beta-thio-galactoside (IPTG) inducible *PtacI* promoter [15] and terminated by the same *TR2* termination [16]. In order to perform measurements on individual cells at low fluorescence levels, we developed a microfluidic flow cytometer ( $\mu$ FACS) based on the work of Fu *et al.* [17]. Using the hypothesis described in the “Calculation of the mean and standard deviation of the plasmid copy number”, the analysis of the first and second moments of the expression of both genes within the whole population allows us to estimate the PCN mean and variance.

## II. MATERIALS AND METHODS

### A. Strains

We modified the Top10 (Clontech) strain which is a derivative of *Escherichia Coli K-12* as follows : Chromosomal copy of *lacI* was replaced by *lacIq1* allele [18] by homologous



recombination using the Counter-Selection BAC Modification Kit from Genebridges [19] (Genebridges GmbH, Dresden, Germany). The *lacIq1* allele represses the *PtacI* promoter a hundred times more efficiently than the wild type *lacI*. Using the same homologous recombination technique, EGFP expression construct was cloned at the opposite of the origin of replication, next to the *pyrF* gene (1339945-1340682 bp).

## B. Expression system

The EGFP expression system was constructed with the *egfp* gene, the *PtacI* promoter [15] and the *TR2* termination sequence [16]. The *egfp* sequence was amplified from the plasmid pEGFP (Clontech). The promoter and the termination were added through two consecutive PCRs using overhanging oligonucleotides. The mOrange [14] expression system was constructed identically. *mOrange* gene was amplified from pRSET-BmOrange [14].

## C. Plasmids

Four plasmids were studied : pZC320 [20] is a mini-F containing a *sopABC* partitioning system, pOU82 is a mini-R1-*par*<sup>-</sup> [21] without a partitioning system, pOAR64 is a mini-R1-*par*<sup>+</sup> [22] with a *parABS* partitioning system and pBR322 is a mini-ColE1. Both mini-R1 plasmids have the same temperature-sensitive origin of replication. In these plasmids, an mRNA coding for the replication initiator RepA can be expressed through a  $\lambda$  promoter. This promoter is strongly inhibited by CI857 which is also constitutively expressed from the plasmid. At high temperature, CI857 is unstable which results in high levels of RepA. The mOrange construct and the plasmids were assembled following standard restriction/ligation protocols. All plasmids were sequenced to verify proper insertion (Eurofins/MWG/Operon). All enzymes used were from Fermentas and all purification kits were from Qiagen.

## D. Microfluidic device

Molds containing a 2 $\mu$ m height motif were made through usual soft lithography techniques with SU8-2002 photoresist (MicroChem) using chrome deposited masks (Dupont photomask). The motif, which was based on Fu *et al.* [17], consists of three channels joined

in a Y-shape junction with branches  $5\mu\text{m}$  in width. Elastomer (RTV615, GE Bayer) was poured into the mold and baked at  $90^\circ\text{C}$  for 2 hours. The cross-linked elastomer was peeled and then drilled to make the reservoirs. The elastomer and the quartz coverslip were placed in a home made plasma cleaner for about 5 seconds. Then, the molded elastomer was sealed to the quartz coverslip. The chip was immediately filled with phosphate buffer saline (PBS,  $\text{pH}=7.4$ ). The bacterial culture of interest was placed in one of the reservoirs and allowed to flow under pressure within channels. Typical flow speed is  $5 \times 10^{-3} \text{ ms}^{-1}$ .

### E. Microscopy and data analysis

The chip was placed on an inverted microscope (IX81, Olympus) and observations were performed using an oil immersion 63x Objective (PlanApo, Zeiss). Bacterial fluorescence was excited with a solid state 488 nm laser (Sapphire 488 HP, Coherent). The laser light wave front was modified using a Spatial Light Modulator (LCR-720 , HOLOEYE Photonics AG) to obtain a homogeneous restricted excitation area. The flow channel was placed within this area. The light intensity was on the order of a few mW. Fluorescence was collected separately on green and orange channels. Light intensity was measured using avalanche photodiodes (SPCM-AQR-15, Perkin Elmer). TTL pulses from avalanche photodiodes were counted by an acquisition board (NI-PCI6602, National Instruments). The experiment was controlled using a custom made LabView interface (National Instruments). Considering the bacterial flow rate, fluorescent proteins are excited on a millisecond time scale. It is far too short to take into account any photobleaching effect. Data were analysed using IGOR 5.1 and Matlab 2006b programs.

### F. Quantitative PCR

Bacteria were grown as described in the next section. The total DNA was extracted from a 1 ml sample of bacteria bearing the various plasmids (Qiagen Qiamap). qPCR was performed on a Light Cycler (Roche Diagnosis) using SYBRgreen as a fluorescent intercalant agent (Qiagen). Oligonucleotides were designed to target 2 genes: *bla* and *pyrF*. All our plasmids contained the *bla* sequence which codes for the ampicillin resistance. The *pyrF* sequence is located at the termini of the chromosome, next to the cloned *egfp* sequence.

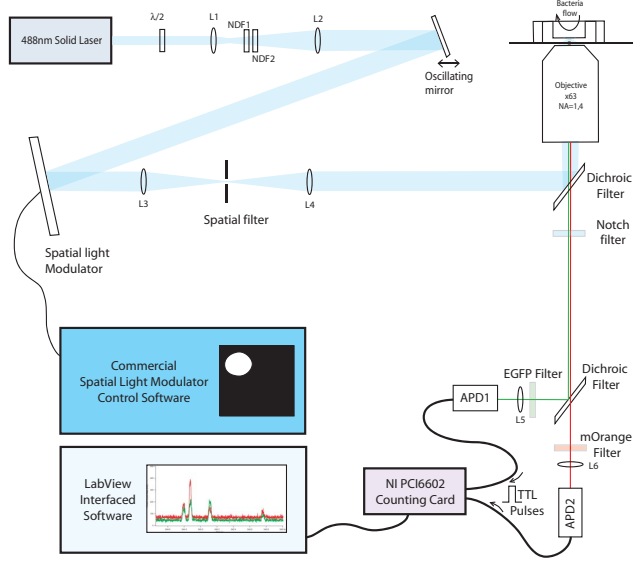


FIG. 1: The experimental set up is built on an inverted microscope. Laser light is shaped to a constant intensity disk of  $\sim 20\mu\text{m}$  diameter in the field of view by a spatial light modulator (SLM). Fluorescent light is collected by the same objective and separated on green and orange channels through a combination of dichroic and emission filters. A sub-region of the field of view (a  $\sim 10\mu\text{m}$  disk) containing the channel through which the bacteria pass is focused on the avalanche photodiodes. TTL pulses from the avalanche photodiodes are fed to a counting acquisition board in a computer. A bacterium passing through the field of view causes an increase in the measured fluorescence intensity.

Five dilutions of each extract were amplified with the 2 sets of oligonucleotides. Data were analysed with LightCycler3 data analysis software.

### G. Bacterial culture and fluorescent protein expression

Fresh overnight cultures from an isolated colony was diluted 500 times in Luria-Bertani (LB) medium supplemented with the necessary antibiotics and grown at the chosen temperature with vigorous shaking (200 RPM) up to an optical density at 600 nm ( $OD_{600}$ ) between 0.20 and 0.25. At this stage, the culture was diluted again by 100 times with fresh medium and grown up to an  $OD_{600} = 0.20 - 0.25$ . IPTG was added to a final concentration of 1 mM and the culture was incubated for another hour. At this point, chloramphenicol was added

to a final concentration of  $170\mu\text{g/ml}$  to stop any additional translation of fluorescent protein. The cells were then washed into PBS with  $170\mu\text{g/ml}$  of chloramphenicol. Bacteria were placed back into the incubator at  $37^\circ\text{C}$  for further maturation of the fluorescent proteins overnight. The pH of the maturation medium is set to 7.4 by PBS. This value is suitable for both mOrange and EGFP stability. From this step on, care was taken to avoid unnecessary exposure to light in order to prevent photobleaching of both fluorescent proteins. The cells were washed again in PBS, filtered through a Millipore syringe filter with a  $5\mu\text{m}$  pore size to remove cell aggregates and finally pipetted in the microfluidic chip. Typical fluorescence measurements required an hour long flow of bacteria for about  $10^4$  detected events.

### III. CALIBRATION AND DATA ANALYSIS

#### A. Equal expression of fluorescent reporters

Our data analysis relies on the fact that both EGFP and mOrange reporters are expressed identically from transcription to protein folding. We tested this assumption by comparing EGFP expression to mOrange expression in the case of a mini-ColE1 plasmid. One mini-ColE1 plasmid was constructed with an EGFP expression system and another mini-ColE1 plasmid was constructed with an mOrange expression system. Chromosomal *lacIq1* strains were transformed with these plasmids. Upon normalization of the mean, measured orange and green distributions superimposed (figure 2). We concluded that both fluorescent protein expressions have the same statistical properties.

#### B. Autofluorescence and fluorescence leakage corrections

In each optical channel, the measured fluorescence can be written as the sum of the fluorescence from the reporter of interest, the autofluorescence of the bacteria and the leakage from the other fluorescent reporter protein. We therefore write:

$$F_G = P_G + A_G + \alpha P_O \quad (1)$$

$$F_O = aP_O + A_O + \beta P_G \quad (2)$$

where  $F_G$  and  $F_O$  are the measured fluorescence in green and orange respectively,  $P_G$  and  $P_O$  are the fluorescence issued respectively from the EGFP and mOrange proteins.  $\alpha$  and

$\beta$  are the leak coefficients.  $a$  is a constant coefficient, which normalizes orange and green excitation and detection efficiencies. It allows us to directly compare  $P_G$  and  $P_O$ .

The parameters  $\alpha$ ,  $\beta$  and  $a$  were estimated using bacteria exhibiting only one type of fluorescence (figure 2).

From equations 1 and 2, moments of  $P_G$  and  $P_O$  can be written as functions of the moments of  $F_G$ ,  $F_O$ ,  $A_G$  and  $A_O$  assuming that the autofluorescence is uncorrelated to the expression of the fluorescent protein. For example, mean values  $\langle P_G \rangle$  and  $\langle P_O \rangle$  are written:

$$\langle P_G \rangle = \frac{a}{a - \alpha\beta} \left\{ (\langle F_G \rangle - \langle A_G \rangle) - \frac{\alpha}{a} (\langle F_O \rangle - \langle A_O \rangle) \right\} \quad (3)$$

$$\langle P_O \rangle = \frac{1}{a - \alpha\beta} \{ (\langle F_O \rangle - \langle A_O \rangle) - \beta (\langle F_G \rangle - \langle A_G \rangle) \} \quad (4)$$

Moments of  $A_G$  and  $A_O$  were determined separately by measurements on non-fluorescent bacterial populations (black population in figure 2).

### C. Calculation of the mean and standard deviation of the plasmid copy number

The mean and standard deviation of the PCN can be calculated rigorously from the measured expression of *egfp* and *mOrange* [23]. For simplicity, we present an approach based on a few basic assumptions. In particular, we suppose that the number of chromosomes and plasmids are kept constant during the induction time and that no cell division occurs. Moreover, we assume that all bacteria have the same cell cycle duration and that ages are uniformly distributed. Again, although these hypotheses are not valid in our experimental setup, the rigorous calculus leads to similar values. Global events such as division or cell age act on both expression from the chromosome and from the plasmid. As we divide mean orange expression by mean green expression (see below), we get rid of most of these global fluctuations. We believe this is the underlying reason why the reductive case leads to values similar to the rigorous calculus.

Inside a bacterium, the green fluorescence intensity  $P_G$  is the sum of the fluorescence of the EGFP expressed from each copy of the chromosome, that is:

$$P_G = \sum_{i=1}^{n_C} P_G^i \quad (5)$$

where  $n_C$  is the number of copies of the chromosome and  $P_G^i$  is the fluorescence of the EGFP proteins expressed from the  $i^{th}$  chromosome copy.

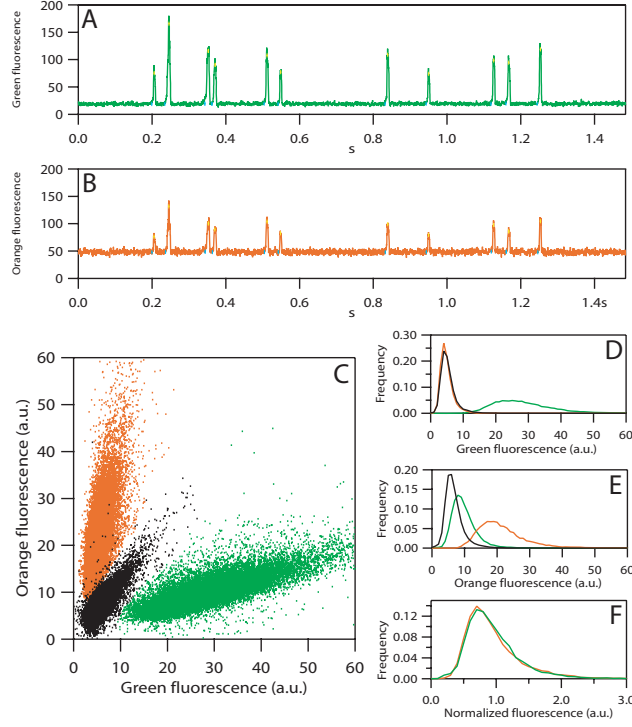


FIG. 2: **(A-B)** Temporal signal from the counting module respectively for the green and for the orange fluorescences. An abrupt increase is perceived when a bacterium passes through the detection field of the photon counting module. The peak and baseline of each event are spotted in blue and black respectively. **(C)** Fluorescence scatter plot of 3 different populations. Bacteria bearing the green expression system on the chromosome without any plasmid is in green. Bacteria with the orange expression system on the same location as the previous strain also without a plasmid is in orange. The original strain bearing no fluorescence is in black. The leakage coefficients  $\alpha$ , and  $\beta$  and the normalization factor  $a$  (see section “Mean and standard deviation of the plasmid copy number calculation”) are deduced from the difference in mean fluorescence of those 3 populations.  $\alpha$  could not be measured as it was too small ( $< 1\%$ ) and  $\beta$  was estimated to  $0.1674 \pm 0.0005$ .  $a$  was calculated to be  $0.58 \pm 0.04$ , which corrects for differences in optical pathways, fluorescent excitation, emission efficiency and differences in expression. **(D-E)** Population histograms of green and orange fluorescence shown in **(C)**. **(F)** Normalized histograms of EGFP and mOrange expressed from a mini-colE1 plasmid.

The same argument applies for the orange fluorescence  $P_O$ . It leads to:

$$P_O = \sum_{i=1}^{n_P} P_O^i \quad (6)$$

where  $n_P$  is the PCN and  $P_O^i$  is the fluorescence of the mOrange proteins expressed from the  $i^{th}$  plasmid copy.

We assume the expression from the chromosome has the same statistical properties as expression from the plasmid. For instance, the mean and standard deviation of the fluorescence of EGFP protein from one copy of the chromosome are equal to the mean and standard deviation of the fluorescence of mOrange protein from one copy of a plasmid. Then, taking the mean over the population, one gets:

$$n = \frac{\langle n_P \rangle}{\langle n_C \rangle} = \frac{\langle P_O \rangle}{\langle P_G \rangle} \quad (7)$$

where  $n$  is the mean PCN per chromosome, which only relies on our measurements and is independent of any other constant. It is important to note that we determine the average plasmid number per chromosome instead of the average plasmid number per cell. We are unable to directly deduce the number of plasmids and chromosomes from the measured green and red fluorescence intensities. This is because of the external noise in gene expression [12] and that the autofluorescence varies from cell to cell (figure 2).

We assume the replication of the plasmid to be independent of the replication of the chromosome and that  $\langle n_C^2 \rangle = \langle n_C \rangle^2$ . The solution for  $\langle n_P^2 \rangle$  then leads to the expression of the variance in PCN:

$$\sigma_{n_P}^2 = \langle n_P \rangle \langle n_C \rangle \left\{ 1 - \frac{\langle n_P \rangle}{\langle n_C \rangle} + \frac{1}{\langle P_O P_G \rangle} \left[ \langle P_O^2 \rangle - \frac{\langle n_P \rangle}{\langle n_C \rangle} \langle P_G^2 \rangle \right] \right\} \quad (8)$$

One can then simply calculate the coefficient of variation or noise in the PCN by dividing the standard deviation by the mean :

$$\eta_{n_P} = \frac{\sigma_{n_P}}{\langle n_P \rangle} \quad (9)$$

Here, the noise only depends on the fluorescence measurement whereas the calculation of the standard deviation needs the mean copy number of chromosome. Because this mean copy number varies a lot with the species and growth conditions [24], we will only discuss the noise in PCN,  $\eta_{n_P}$ .

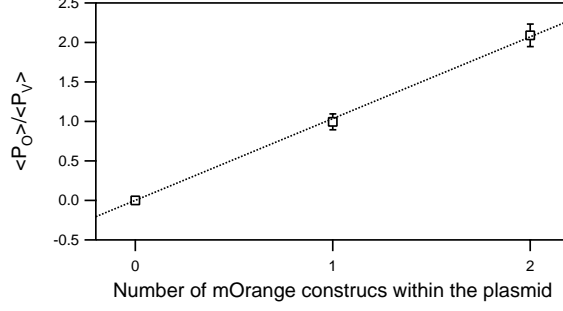


FIG. 3: Mean expression from the mini-F plasmid normalized to the expression from the chromosome as a function of the number of *mOrange* constructs in the mini-F plasmid. The line is a linear fit with a slope of  $1.03 \pm 0.02$ .

## IV. RESULTS AND DISCUSSION

### A. Expression is linear with gene copy number

In order to verify the linearity of expression with gene copy number, two *mOrange* expression systems were cloned in the mini-F plasmid. Mean expression from the plasmid normalized to the expression from the chromosome scaled linearly to the number of *mOrange* constructs within the plasmid (figure 3). One might expect the expression from a single copy to lower with an increasing copy number due to limited resources within the bacterium. However, this is not the case. The green fluorescence did not change significantly (table 1) whether the bacteria contained no plasmid or any plasmid of our series expressing *mOrange*. We therefore assume that the mean expression over a population is linear with respect to the gene copy number.

### B. Determination of the plasmid copy number mean and noise

The mean PCN per chromosome for each plasmid was determined using formula 7. Results at 37°C are summarized in table 1. Guyer *et al* [25] measured a similar value for the mini-F. While no PCN measurement of our mini-R1 was reported, equivalent values were found for other derivatives of the R1 plasmid [26]. Finally, ColE1 based plasmids are generally accepted to have a PCN around 30 [27]. We measured a mean PCN 3 times higher for this plasmid. This discrepancy may result from measurement techniques as discussed



TABLE I: Expression and deduced PCN mean and noise values at 37°C. Each value is the mean of at least 5 experiments. The errors are the standard deviation between the different experiments.  $P_O^*$  is the *mOrange* expression  $\langle P_O \rangle$  divided by the mean *mOrange* expression from a strain containing the mOrange construct on the chromosome (figure 2).  $n_{qpcr}$  is the measured mean PCN from the qPCR experiments.

	mini-F	mini-R1- <i>par</i> <sup>-</sup>	mini-R1- <i>par</i> <sup>+</sup>	mini-ColE1
$\langle P_G \rangle$	$27.1 \pm 0.8$	$28.5 \pm 0.8$	$26.5 \pm 0.4$	$25.7 \pm 1.9$
$\langle P_O \rangle$	$27.0 \pm 0.9$	$244 \pm 16$	$173 \pm 8$	$2167 \pm 89$
$n = \langle P_O \rangle / \langle P_G \rangle$	$1.0 \pm 0.04$	$7.8 \pm 0.2$	$6.5 \pm 0.3$	$95 \pm 2$
$\langle P_O^* \rangle$	$1.1 \pm 0.2$	$10 \pm 2$	$7 \pm 1$	$88 \pm 16$
$n_{qpcr}$	$0.5 \pm 0.3$	$3 \pm 1$	$4 \pm 1$	$25 \pm 5$
$\eta_{n_P} (\times 10^{-2})$	$46 \pm 2$	$34 \pm 2$	$29.2 \pm 0.6$	$25 \pm 3$

below.

Expression from a plasmid seemed equivalent to the expression from the chromosome as  $\langle P_O^* \rangle$  showed similar values to  $n$ .  $n$  was chosen over  $\langle P_O^* \rangle$  as a measurement of the PCN per chromosome because it integrates a reference within each experiment. PCN estimated by qPCR are approximately a factor of 2 lower than values from fluorescence measurement for the mini-F and R1 plasmids, which increases to 4 times for the mini-colE1. Still, we are confident that our measurement is correct. The DNA extraction kit we used is specifically designed to extract chromosomal DNA. We tested that the extraction columns are less efficient for lighter DNA fragments while loaded with GeneRuler 1kb DNA ladder (Fermentas). Therefore, we strongly suspect the DNA extraction process prior to the qPCR to favor high molecular mass. Finally, our bacterial preparation does not involve an extraction process or labeling efficiency. Since chromosome expression does not change for each plasmid, systematic over-estimation of the mean PCN is unlikely. This makes our PCN measurements more reliable.

The major contribution of our work is the estimation of PCN noise. The mini-F showed a noise of about 50% which is consistent with having between 1 and 2 plasmid copies within each cell. Both mini-R1 had noise figures very close to simulations from Paulsson *et al.* [8]. Concerning ColE1, Goss et Peccoud [28] evaluate by numerical simulations PCN noise of

about 20% which fairly matches our  $25 \pm 3\%$ .

### C. *par* system reduces both mean and noise of plasmid copy number

We observed a lower expression level from the mini-R1-*par*<sup>+</sup> than from the mini-R1-*par*<sup>-</sup>. One possible reason could be a gene silencing effect caused by the *par* system. Indeed, it is known that the *par* system induces silencing of the neighboring genes [29]. This phenomenon is thought to result from ParB proteins binding to the *parS* sequence and then spreading along the DNA. Our mini-R1-*par*<sup>+</sup> contains 5 iteron sequences which are RepA protein binding sites. RepA binding prevents the polymerization of ParB on the DNA. It has been shown that genes separated from the *parS* by this same sequence display no silencing effect [29]. Therefore, we concluded that lower expression from the mini-R1-*par*<sup>+</sup> was due to a lower copy number.

The lowering of the plasmid copy number can occur for two non-exclusive reasons. The first one is not linked to the partitioning system but to the iteron sequence. The RepA protein needed to suppress silencing initiates replication of the mini-R1 plasmids. Hence, a sequestering of RepA by the iteron repeat sequence would lead to lower replication frequency and result in a reduced PCN. The second reason is directly caused by the *par* system. The partitioning system is known to reduce loss probability [30]. Typically, when the plasmid is lost, a bacterium containing a “high” copy number is born. If selection for the plasmid bearing bacteria occurs, the resulting population will have a greater mean PCN. Therefore, the lack of partitioning system increases the plasmid loss rate, resulting in an increase in mean PCN.

The mini-R1-*par*<sup>+</sup> also showed a reduction in PCN noise relative to the mini-R1-*par*<sup>-</sup>. The noise in PCN is mainly due to two phenomena: At the cell division, plasmids are not distributed equally between the daughter cells. This occurs either because plasmids are randomly distributed or simply because there is an uneven number of plasmids to distribute. Second, even in bacteria bearing the same number of plasmids at birth, the number of replication events and their timing can be different, which leads to a different PCN in the end. Both mini-R1 have the same replication system. As expected, we deduced it is the *par* system that reduces the noise in PCN.

#### D. Temperature dependance of the Plasmid Copy Number and its variability

The mini-R1 PCN regulation networks in this study were designed to be temperature sensitive [26]. At temperatures below 37°C, these plasmids behave as regular low copy R1 plasmids whereas at temperatures higher than 37°C, *repA* expression is increased and results in increased or even uncontrolled plasmid replication [26]. This increase in mean PCN was estimated with single cell resistance to ampicillin [31].

We measured both mini-R1 PCN between 30°C and 39°C (figure 4). PCN slowly increased between 30°C and 37°C by 46% for the mini-R1-*par*<sup>+</sup> and by 16% for the mini-R1-*par*<sup>-</sup>. PCN increased abruptly at 39°C. This abrupt response was also reported by Uhlin *and al.* [31]. The *par* system lowered the PCN at all temperatures, although the difference is less significant at 39°C. The coefficient of variation or noise was nearly constant from 30 to 37°C, then had a 2-fold increase at 39°C (figure 4). Furthermore, there was no visible effect of the *par* system at 39°C on PCN noise. These results indicate that the *par* system is less effective as the PCN is increased.

All the plasmids had an increasing PCN with temperature increase (figure 4). This shows that temperature and duplication time are important parameters when considering the replication regulation process. It is even more important if the replication frequency is low : The mini-F PCN jumped from 0.5 to 2 and the mini-ColE1 PCN evolved from 76 to 100. An interesting feature is that the “relaxed” plasmid ColE1 has the lowest noise in PCN (23 – 25%) whereas the stringent mini-F plasmid has a relatively high noise in PCN (37 – 55%) (see figure 4). For the mini-F, the tendency is that an increase in copy number is associated with a diminution of noise. However, for the other plasmids, noise was nearly constant. This indicates that PCN noise is inherent to the regulation processes coded on each plasmid.

#### E. Plasmid loss rate

We also tested the stability of the studied plasmids on  $\sim 100$  generations. Every plasmid that we used coded for the identical ampicillin resistance. Bacteria bearing plasmids were allowed to grow in a medium without ampicillin. A single colony was inoculated in LB medium and the culture was grown overnight at 37°C. Each overnight culture was diluted

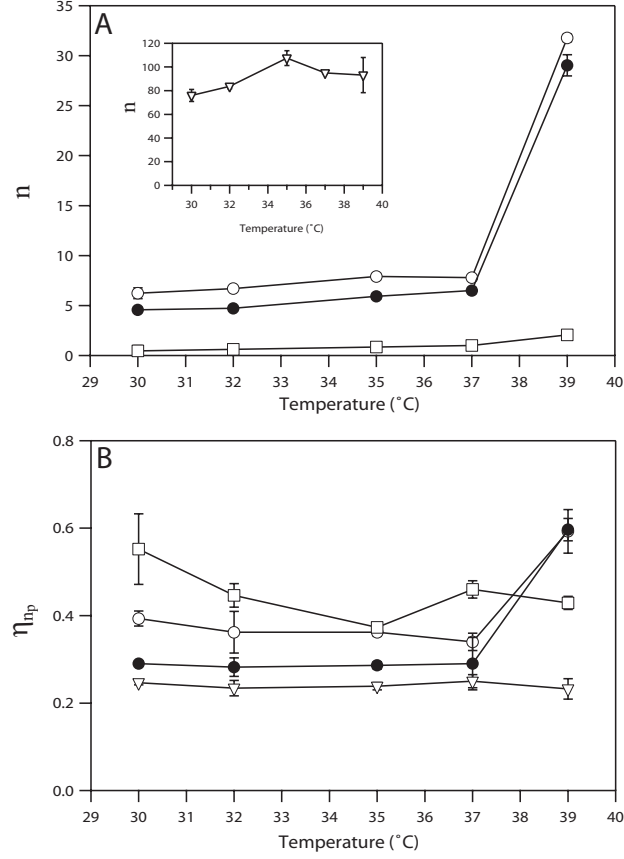


FIG. 4: Mean (A) and noise (B) in PCN for the mini-F (squares), the mini-ColE1 (triangle, inset), the mini-R1-*par*<sup>-</sup> (open circle) and the mini-R1-*par*<sup>+</sup> (closed circle) as a function of temperature. The error bars are standard deviations obtained from multiple experiments. Error bars were not plotted whenever they were smaller than the data point size.

500 times. As the bacteria double their number at each generation, it takes  $\log_2(500) \sim 9$  generations to overcome the dilution. Each strain was re-cultured every evening for ten days. We stored a 1 ml sample of each overnight culture in glycerol at  $-20^\circ\text{C}$ . Prior to fluorescent measurements, the sample was thawed and cultured following the protocol described in the section “Bacterial culture and fluorescent protein expression”. A bacterium exhibiting only EGFP fluorescence was considered cured from the plasmid. Experiments were performed on bacteria at generation 54 (5 overnight cultures and induction process) and 99 (10 overnight and induction process).

The mini-F and the mini-R1-*par*<sup>+</sup> both contained a partitioning system and proved to be completely stable. The mini-ColE1 showed no plasmid loss on the timescale of our

experiment. This is most likely a consequence of its high PCN. The only plasmid that proved unstable was the mini-R1-*par*<sup>-</sup>. We extracted the proportion  $\Psi^+(g)$  at generation  $g$  of bacteria harbouring plasmids from our fluorescence measurements (see figure 5). As a control experiment, this proportion was also evaluated by replica plating and proved to be consistent with fluorescent measurements.

Boe and Rasmussen [32, 33] proposed two methods to calculate the plasmid loss rate. The first one assumes that the bacteria bearing plasmids and those without plasmids have the same division time. The proportion of bacteria cured from the plasmid  $\Psi^-$  increases through the generations solely because of the plasmid loss rate  $\theta_1$  which should be the same at any generation. However, our estimations of  $\theta_1$  from different generations gave different values indicating that the second method should be used. The second method postulates a growth inhibition effect on plasmids bearing bacteria. This results in a different division time for bacteria bearing plasmids ( $T^+$ ) and for cured bacteria ( $T^-$ ) with  $T^+ > T^-$ . Here,  $\Psi^-$  rises because of the plasmid loss rate  $\theta_2$  but also because the cured bacteria grow faster. Our data analysis gave the same  $\theta_2$  for the different generations ( $\theta_2 \simeq 0.53 - 0.57\%$  at  $37^\circ\text{C}$ ). Moreover, we were able to evaluate the growth inhibition effect as  $T^+/T^- \simeq 1.066$  which corresponds to a difference in division time of about 2 minutes.

## V. CONCLUSION

We designed an experiment that allowed us to measure the mean and the standard deviation of PCN for a set of plasmids. This is to our knowledge, the first time the noise in PCN has been measured. We observed that the partitioning system decreased the PCN while also diminishing noise in PCN. We found that the mean PCN always varied with temperature regardless of the replication origin. Moreover, PCN noise showed either no or a relatively smooth change with temperature with the exception of the temperature sensitive plasmids, for which noise increased abruptly at  $39^\circ\text{C}$ . Another feature is that random partitioning at high plasmid copy numbers induces less noise in PCN than low plasmid copy numbers with a partitioning system. Finally, we distinguished cured bacteria from plasmid bearing bacteria. This allowed us to measure the plasmid loss rate of the mini-R1-*par*<sup>-</sup> and the growth inhibition effect due to this plasmid.

Finally, it is known that the mean PCN of some plasmids can be changed by applying

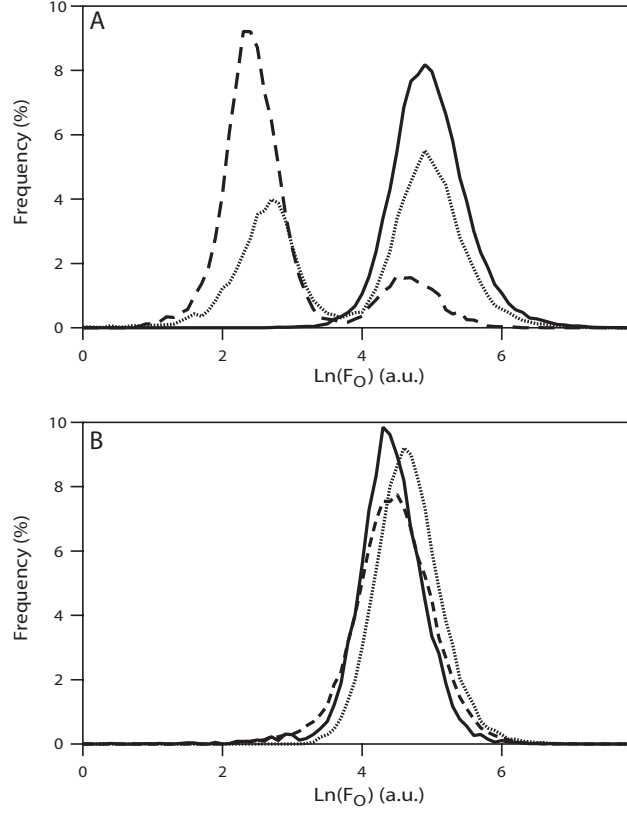


FIG. 5: Histograms of the orange fluorescence ( $F_O$ ) in logarithmic scale for the mini-R1- $par^-$  (**A**) and the mini-R1- $par^+$  (**B**) at different generation times. In black (plain), the plasmids are in the presence of pressure selection. The red (dotted) and blue (dashed) histograms of the population are the fluorescence distributions after 54 and 99 generations in the medium without antibiotics, respectively.

a different selection pressure such as antibiotic concentration [34]. Moreover, Spudich and Koshland suggested that variability in chemotactic response could be a source of adaptability on the population level [35]. With our experimental set up and analysis, we hope to address the question of whether the variability in PCN is a source of adaptability.

## Acknowledgments

J.W.N. acknowledges financial support from the Ministère de la Recherche and CNRS. This work was supported by the grant n°05-BLAN-0026-01 from the ANR.

---

- [1] M. L. Holmes, F. Pfeifer, and M. L. Dyall-Smith, *Gene* **153**, 117 (1995).
- [2] C. Dahlberg, M. Bergstrom, M. Andreassen, B. Christensen, S. Molin, and M. Hermansson, *Mol Biol Evol* **15**, 385 (1998).
- [3] K. Gerdes, L. K. Poulsen, T. Thisted, A. K. Nielsen, J. Martinussen, and P. H. Andreassen, *New Biol* **2**, 946 (1990).
- [4] G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas, *Microbiol Mol Biol Rev* **62**, 434 (1998).
- [5] J. Tomizawa and T. Itoh, *Proc Natl Acad Sci U S A* **78**, 6096 (1981).
- [6] K. Nordström and E. G. Wagner, *Trends Biochem Sci* **19**, 294 (1994).
- [7] P. J. Goss and J. Peccoud, *Proc Natl Acad Sci U S A* **95**, 6750 (1998).
- [8] J. Paulsson and M. Ehrenberg, *Q Rev Biophys* **34**, 1 (2001).
- [9] C. Lee, J. Kim, S. G. Shin, and S. Hwang, *J Biotechnol* **123**, 273 (2006).
- [10] A. S. Belmont, *Trends Cell Biol* **11**, 250 (2001).
- [11] J. Pogliano, T. Q. Ho, Z. Zhong, and D. R. Helinski, *Proc Natl Acad Sci U S A* **98**, 4486 (2001).
- [12] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).
- [13] R. Y. Tsien, *Annu Rev Biochem* **67**, 509 (1998).
- [14] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien, *Nat Biotechnol* **22**, 1567 (2004).
- [15] H. A. de Boer, L. J. Comstock, and M. Vasser, *Proc Natl Acad Sci U S A* **80**, 21 (1983).
- [16] I. Gusarov and E. Nudler, *Mol Cell* **3**, 495 (1999).
- [17] A. Y. Fu, C. Spence, A. Scherer, F. H. Arnold, and S. R. Quake, *Nat Biotechnol* **17**, 1109 (1999).
- [18] C. B. Glascock and M. J. Weickert, *Gene* **223**, 221 (1998).
- [19] *Counter-selection BAC modification kit (Advanced BAC Modification kit)*, Gene Bridges

- (2004).
- [20] J. Shi and D. P. Biek, *Gene* **164**, 55 (1995).
  - [21] K. Gerdes, J. E. Larsen, and S. Molin, *J Bacteriol* **161**, 292 (1985).
  - [22] O. Rodionov and M. Yarmolinsky, *Mol Microbiol* **52**, 1215 (2004).
  - [23] S. Ghozzi (2009), in preparation.
  - [24] F. C. Neidhardt, *Escherichia coli and Salmonella , Cellular and molecular biology*, asm press ed. (1996).
  - [25] M. S. Guyer, R. R. Reed, J. A. Steitz, and K. B. Low, *Cold Spring Harb Symp Quant Biol* **45 Pt 1**, 135 (1981).
  - [26] J. E. Larsen, K. Gerdes, J. Light, and S. Molin, *Gene* **28**, 45 (1984).
  - [27] Sambrook and Russel, *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd ed. (2001).
  - [28] P. J. Goss and J. Peccoud, *Pac Symp Biocomput* pp. 65–76 (1999).
  - [29] O. Rodionov, M. Lobočka, and M. Yarmolinsky, *Science* **283**, 546 (1999).
  - [30] S. Austin, S. Friedman, and D. Ludtke3, *J Bacteriol* **168**, 1010 (1986).
  - [31] B. E. Uhlin and K. Nordström, *J Bacteriol* **124**, 641 (1975).
  - [32] L. Boe, *Plasmid* **36**, 161 (1996).
  - [33] L. Boe and K. V. Rasmussen, *Plasmid* **36**, 153 (1996).
  - [34] A. Løbner-Olesen, *EMBO J* **18**, 1712 (1999).
  - [35] J. L. Spudich and D. E. Koshland, *Nature* **262**, 467 (1976).



# Bibliographie

- [1] F. Crick, "Central dogma of molecular biology.," *Nature*, vol. 227, pp. 561–563, Aug 1970.
- [2] I. Gusarov and E. Nudler, "The mechanism of intrinsic transcription termination.," *Mol Cell*, vol. 3, pp. 495–504, Apr 1999.
- [3] J. M. G. Vilar, C. C. Guet, and S. Leibler, "[Modeling network dynamics : the lac operon, a case study.](#)," *J Cell Biol*, vol. 161, pp. 471–476, May 2003.
- [4] R. M. Schaaper, "Base selection, proofreading, and mismatch repair during dna replication in escherichia coli.," *J Biol Chem*, vol. 268, pp. 23762–23765, Nov 1993.
- [5] A. F. Bennett and R. E. Lenski, "[An experimental test of evolutionary trade-offs during temperature adaptation.](#)," *Proc Natl Acad Sci U S A*, May 2007.
- [6] J. L. Spudich and D. E. Koshland, "Non-genetic individuality : chance in the single cell.," *Nature*, vol. 262, pp. 467–471, Aug 1976.
- [7] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression.," *Proc Natl Acad Sci U S A*, vol. 94, pp. 814–819, Feb 1997.
- [8] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, "[Regulation of noise in the expression of a single gene.](#)," *Nat Genet*, vol. 31, pp. 69–73, May 2002. utilise B. Subtilis. fait pousser à 37°C dans LB overnight. dilue x'???' fois. Ensuite induit à l'IPTG au moins pendant 5h. D'autre part, il dit qu'il prélève les bactéries à OD 1.0... J'atteins l'OD1 au bout de 250-300 minutes—>Il induit dès le départ donc...
- [9] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "[Stochastic gene expression in a single cell.](#)," *Science*, vol. 297, pp. 1183–1186, Aug 2002. Fait pousser à 32°C overnight, ensuite innocule 1/100e jusqu'à OD 0.2-0.3. ne précise pas le moment de l'induction... qu'est-ce que cela veut dire???
- [10] M. Thattai and A. van Oudenaarden, "[Intrinsic noise in gene regulatory networks.](#)," *Proc Natl Acad Sci U S A*, vol. 98, pp. 8614–8619, Jul 2001.
- [11] D. Adalsteinsson, D. McMillen, and T. C. Elston, "[Biochemical network stochastic simulator \(bionets\) : software for stochastic modeling of biochemical networks.](#)," *BMC Bioinformatics*, vol. 5, p. 24, Mar 2004.
- [12] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "[Network motifs in the transcriptional regulation network of escherichia coli.](#)," *Nat Genet*, vol. 31, pp. 64–68, May 2002.
- [13] K. R. Hopper, "[Risk-spreading and bet-hedging in insect population biology.](#)," *Annu Rev Entomol*, vol. 44, pp. 535–560, 1999.
- [14] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, "[Bacterial persistence as a phenotypic switch.](#)," *Science*, vol. 305, pp. 1622–1625, Sep 2004.
- [15] E. Kussell and S. Leibler, "[Phenotypic diversity, population growth, and information in fluctuating environments.](#)," *Science*, vol. 309, pp. 2075–2078, Sep 2005.
- [16] E. Kussell, R. Kishony, N. Q. Balaban, and S. Leibler, "[Bacterial persistence : a model of survival in changing environments.](#)," *Genetics*, vol. 169, pp. 1807–1814, Apr 2005.
- [17] L. Snyder and W. Champness, *Molecular genetics of bacteria*. Washington D.C : ASM Press, cop., 2nd ed. ed., 2003.
- [18] T. A. Baker and S. H. Wickner, "[Genetics and enzymology of dna replication in escherichia coli.](#)," *Annu Rev Genet*, vol. 26, pp. 447–477, 1992.
- [19] A. Løbner-Olesen, K. Skarstad, F. G. Hansen, K. von Meyenburg, and E. Boye, "The dnaa protein determines the initiation mass of escherichia coli k-12.," *Cell*, vol. 57, pp. 881–889, Jun 1989.
- [20] Q. P. Wang and J. M. Kaguni, "Transcriptional repression of the dnaa gene of escherichia coli by dnaa protein.," *Mol Gen Genet*, vol. 209, pp. 518–525, Oct 1987.
- [21] E. Crooke, "Regulation of chromosomal replication in e. coli : sequestration and beyond.," *Cell*, vol. 82, pp. 877–880, Sep 1995.
- [22] S. Jun and B. Mulder, "[Entropy-driven spatial organization of highly confined polymers : lessons for the bacterial chromosome.](#)," *Proc Natl Acad Sci U S A*, vol. 103, pp. 12388–12393, Aug 2006.
- [23] S. Hiraga, "[Dynamic localization of bacterial and plasmid chromosomes.](#)," *Annu Rev Genet*, vol. 34, pp. 21–59, 2000.
- [24] G. del Solar and M. Espinosa, "Plasmid copy number control : an ever-growing story.," *Mol Microbiol*, vol. 37, pp. 492–500, Aug 2000.
- [25] J. Lederberg, L. L. Cavalli, and E. M. Lederberg, "Sex compatibility in escherichia coli.," *Genetics*, vol. 37, pp. 720–730, Nov 1952.

- [26] J. Shi and D. P. Biek, "A versatile low-copy-number cloning vector derived from plasmid f.," *Gene*, vol. 164, pp. 55–58, Oct 1995.
- [27] H. Uga, F. Matsunaga, and C. Wada, "Regulation of dna replication by iterons : an interaction between the ori2 and incc regions mediated by repe-bound iterons inhibits dna replication of mini-f plasmid in escherichia coli.," *EMBO J*, vol. 18, pp. 3856–3867, Jul 1999.
- [28] G. E. Lim, A. I. Derman, and J. Pogliano, "Bacterial dna segregation by dynamic sopa polymers.," *Proc Natl Acad Sci U S A*, vol. 102, pp. 17658–17663, Dec 2005.
- [29] T. Watanabe and T. Fukasawa, "Episome-mediated transfer of drug resistance in enterobacteriaceae. i. transfer of resistance factors by conjugation.," *J Bacteriol*, vol. 81, pp. 669–678, May 1961.
- [30] H. Masai, Y. Kaziro, and K. Arai, "Definition of orir, the minimum dna segment essential for initiation of r1 plasmid replication in vitro.," *Proc Natl Acad Sci U S A*, vol. 80, pp. 6814–6818, Nov 1983.
- [31] L. Koppes and K. Nordström, "Insertion of an r1 plasmid into the origin of replication of the e. coli chromosome : random timing of replication of the hybrid chromosome.," *Cell*, vol. 44, pp. 117–124, Jan 1986.
- [32] D. D. Womble and R. H. Rownd, "Regulation of incfii plasmid dna replication. a quantitative model for control of plasmid nr1 replication in the bacterial cell division cycle.," *J Mol Biol*, vol. 192, pp. 529–547, Dec 1986.
- [33] K. Nordström and E. G. Wagner, "Kinetic aspects of control of plasmid replication by antisense rna.," *Trends Biochem Sci*, vol. 19, pp. 294–300, Jul 1994.
- [34] M. Ehrenberg and A. Sverredal, "A model for copy number control of the plasmid r1.," *J Mol Biol*, vol. 246, pp. 472–485, Mar 1995.
- [35] O. Rodionov, M. Lobočka, and M. Yarmolinsky, "Silencing of genes flanking the p1 plasmid centromere.," *Science*, vol. 283, pp. 546–549, Jan 1999.
- [36] M. Brenner and J. Tomizawa, "Quantitation of cole1-encoded replication elements.," *Proc Natl Acad Sci U S A*, vol. 88, pp. 405–409, Jan 1991.
- [37] T. Atlung, B. B. Christensen, and F. G. Hansen, "Role of the rom protein in copy number control of plasmid pbr322 at different growth rates in escherichia coli k-12.," *Plasmid*, vol. 41, pp. 110–119, Mar 1999.
- [38] S. Lin-Chao and H. Bremer, "Effect of the bacterial growth rate on replication control of plasmid pbr322 in escherichia coli.," *Mol Gen Genet*, vol. 203, pp. 143–149, Apr 1986.
- [39] P. J. Goss and J. Peccoud, "Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets.," *Proc Natl Acad Sci U S A*, vol. 95, pp. 6750–6755, Jun 1998.
- [40] P. J. Goss and J. Peccoud, "Analysis of the stabilizing effect of rom on the genetic network controlling cole1 plasmid replication.," *Pac Symp Biocomput*, pp. 65–76, 1999.
- [41] M. Ehrenberg, "Hypothesis : hypersensitive plasmid copy number control for cole1.," *Biophys J*, vol. 70, pp. 135–145, Jan 1996.
- [42] J. Paulsson and M. Ehrenberg, "Trade-off between segregational stability and metabolic burden : a mathematical model of plasmid cole1 replication control.," *J Mol Biol*, vol. 279, pp. 73–88, May 1998.
- [43] J. Paulsson, K. Nordström, and M. Ehrenberg, "Requirements for rapid plasmid cole1 copy number adjustments : a mathematical model of inhibition modes and rna turnover rates.," *Plasmid*, vol. 39, no. 3, pp. 215–234, 1998.
- [44] J. Paulsson and M. Ehrenberg, "Noise in a minimal regulatory network : plasmid copy number control.," *Q Rev Biophys*, vol. 34, pp. 1–59, Feb 2001.
- [45] A. S. Belmont, "Visualizing chromosome dynamics with gfp.," *Trends Cell Biol*, vol. 11, pp. 250–257, Jun 2001.
- [46] J. Pogliano, T. Q. Ho, Z. Zhong, and D. R. Helinski, "Multicopy plasmids are clustered and localized in escherichia coli.," *Proc Natl Acad Sci U S A*, vol. 98, pp. 4486–4491, Apr 2001.
- [47] H. A. de Boer, L. J. Comstock, and M. Vasser, "The tac promoter : a functional hybrid derived from the trp and lac promoters.," *Proc Natl Acad Sci U S A*, vol. 80, pp. 21–25, Jan 1983.
- [48] C. B. Glascock and M. J. Weickert, "Using chromosomal lacIq1 to control expression of genes on high-copy-number plasmids in escherichia coli.," *Gene*, vol. 223, pp. 221–231, Nov 1998.
- [49] R. L. Vellanoweth and J. C. Rabinowitz, "The influence of ribosome-binding-site elements on translational efficiency in bacillus subtilis and escherichia coli in vivo.," *Mol Microbiol*, vol. 6, pp. 1105–1114, May 1992.
- [50] R. Y. Tsien, "The green fluorescent protein.," *Annu Rev Biochem*, vol. 67, pp. 509–544, 1998.
- [51] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien, "Improved monomeric red, orange and yellow fluorescent proteins derived from discosoma sp. red fluorescent protein.," *Nat Biotechnol*, vol. 22, pp. 1567–1572, Dec 2004.
- [52] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki, "A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.," *Nat Biotechnol*, vol. 20, pp. 87–90, Jan 2002.
- [53] M. Zuker, D. Mathews, and D. Turner, "Algorithms and thermodynamics for RNA secondary structure prediction : A practical guide," in *In RNA Biochemistry and Biotechnology*, J. Barciszewski & B.F.C. Clark, eds., NATO ASI Series, Kluwer Academic Publishers, 1999.
- [54] Gene Bridges, *Counter-selection BAC modification kit (Advanced BAC Modification kit)*, APRIL 2004.
- [55] H. B. Steen, "Flow cytometry of bacteria : glimpses from the past with a view to the future.," *J Microbiol Methods*, vol. 42, pp. 65–74, Sep 2000.
- [56] S. K. Sia and G. M. Whitesides, "Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies.," *ELECTROPHORESIS*, vol. 24, no. 21, pp. 3563–3576, 2003.
- [57] M. A. McClain, C. T. Culbertson, S. C. Jacobson, and J. M. Ramsey, "Flow cytometry of escherichia coli on microfluidic devices.," *Anal Chem*, vol. 73, pp. 5334–5338, Nov 2001.
- [58] A. Y. Fu, C. Spence, A. Scherer, F. H. Arnold, and S. R. Quake, "A microfabricated fluorescence-activated cell sorter.," *Nat Biotechnol*, vol. 17, pp. 1109–1111, Nov 1999.

- [59] T. Bayraktar and S. B. Pidugu, "Characterization of liquid flows in microfluidic systems," *INTERNATIONAL JOURNAL OF HEAT AND MASS TRANSFER*, vol. 49, pp. 815–824, Mar. 2006.
- [60] P. K. Dasgupta and S. Liu, "Electroosmosis : A reliable fluid propulsion system for flow injection analysis," *Anal. Chem.*, vol. 66, no. 11, pp. 1792–1798, 1994.
- [61] Y. Xia and G. M. Whitesides, "Soft lithography," *Annu. Rev. Mater. Sci.*, vol. 28, pp. 153–84, 1998.
- [62] H. Bremer and P. P. Dennis, *Escherichia coli and Salmonella*, vol. 2, ch. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate, pp. 1553–1569. ASM Press, 2nd ed., 1996.
- [63] C. Lee, J. Kim, S. G. Shin, and S. Hwang, "Absolute and relative qpcr quantification of plasmid copy number in escherichia coli," *J Biotechnol*, vol. 123, pp. 273–280, May 2006.
- [64] M. S. Guyer, R. R. Reed, J. A. Steitz, and K. B. Low, "Identification of a sex-factor-affinity site in e. coli as gamma delta.," *Cold Spring Harb Symp Quant Biol*, vol. 45 Pt 1, pp. 135–140, 1981.
- [65] P. S. Swain, M. B. Elowitz, and E. D. Siggia, "Intrinsic and extrinsic contributions to stochasticity in gene expression.," *Proc Natl Acad Sci U S A*, vol. 99, pp. 12795–12800, Oct 2002.
- [66] L. Boe, "Estimation of plasmid loss rates in bacterial populations with a reference to the reproducibility of stability experiments.," *Plasmid*, vol. 36, pp. 161–167, Nov 1996.
- [67] L. Boe and K. V. Rasmussen, "Suggestions as to quantitative measurements of plasmid loss.," *Plasmid*, vol. 36, pp. 153–159, Nov 1996.
- [68] A. Jaffé, T. Ogura, and S. Hiraga, "Effects of the ccd function of the f plasmid on bacterial growth.," *J Bacteriol*, vol. 163, pp. 841–849, Sep 1985.
- [69] J. Jasiecki and G. Wegrzyn, "Transcription start sites in the promoter region of the escherichia coli pcnb (plasmid copy number) gene coding for poly(a) polymerase i.," *Plasmid*, vol. 55, pp. 169–172, Mar 2006.
- [70] A. Løbner-Olesen, "Distribution of minichromosomes in individual escherichia coli cells : implications for replication control.," *EMBO J*, vol. 18, pp. 1712–1721, Mar 1999.
- [71] N. C. Shaner, P. A. Steinbach, and R. Y. Tsien, "A guide to choosing fluorescent proteins.," *Nat Methods*, vol. 2, pp. 905–909, Dec 2005.
- [72] B. J. Bevis and B. S. Glick, "Rapidly maturing variants of the discosoma red fluorescent protein (dsred).," *Nat Biotechnol*, vol. 20, pp. 83–87, Jan 2002.
- [73] Sambrook and Russel, *Molecular Cloning, a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition ed., 2001.